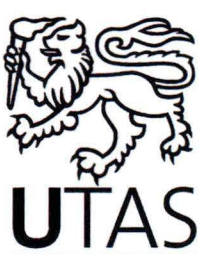
STATISTICAL INFERENCE FOR MOVEMENT BEHAVIOUR USING ANIMAL TRACKING DATA

by

Robin Barbara Thomson, B.Sc. (Zool), B.Sc. Hons (Zool), M.Sc.(Appl. Maths), University of Cape Town

Submitted in fulfilment of the requirements
for the Degree of Doctor of Philosophy

Department of Mathematics
University of Tasmania
October, 2008



I declare that this thesis contains no material which has been accepted for a degree or diploma by the University or any other institution, except by way of background information and duly acknowledged in the thesis, and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due acknowledgement is made in the text of the thesis.

Signed: _____
Robin Barbara Thomson

Date: 10 Oct 2008

Signed: _____
Robin Barbara Thomson

Date: 1 Oct 2008

ABSTRACT

Satellite tracking provides an opportunity to learn about how animals choose to move and about the covariates of movement. Quantitative methodology for this problem has lagged behind the remote sensing technology that provides both animal tracks and covariate information. A statistical framework capable of providing appropriate hypothesis testing has to couple very different types of data: highly autocorrelated time series of observed locations (the track), with 2-dimensional maps of covariate data. In addition, animals respond to internal motivations, representable only as theorised motivations. Behaviour is likely to be highly complex and to be only approximately understood so that process error cannot be ignored. Observation error should be accounted for separately from process error because longitude is typically more difficult to estimate than latitude, and because estimates of observation error are sometimes available. State space models account separately for observation and process errors, and model the serial correlation inherent in tracks. State space models offer great flexibility, nevertheless, the means of incorporating diverse movement behaviours and covariate information is not immediately clear. Traditionally, these models require that time series be equally spaced in time (seldom the case with observed tracks) and they have presented substantial difficulties in inference.

This thesis presents a flexible Bayesian state space modelling framework suitable for application to tracks. Movement behaviour is incorporated through advection fields that represent movement hypotheses. These are calculated using theories regarding movement behaviour, possibly coupled with covariate information. The deviance information criterion DIC measures the weight given by observed track data to alternative proposed hypotheses regarding movement behaviour. In simulation, DIC successfully discriminated the advection fields, and therefore the movement hypotheses, used to simulate track data. DIC is less sensitive than the Bayes factor is, to the priors, an advantage in a field in which little prior information is available. Markov chain Monte Carlo sampling successfully facilitated nonlinear, non-Gaussian model forms while avoiding the inference problems encountered by practitioners of sequential importance sampling. Latent locations are estimated, allowing realistic, nonlinear path estimation. Inertia, a tendency for directional persistence, is incorporated. The Bayesian approach allows the incorporation of prior information and eases inference. Temporal shifts in behaviour are also modelled. The method is demonstrated in practice, using satellite tracks from white sharks in Australia. The problem of modelled animals becoming trapped in semi-enclosed areas and stepping across narrow barriers is discussed and a proposed solution, using Laplace's equation to provide advective flow around obstacles, is demonstrated.

ACKNOWLEDGEMENTS

Many people helped me to complete this thesis. My grateful thanks go to my supervisors - to Simon Wotherspoon for patient clear explanations, good ideas, and for teaching me lots of new words and a new way of thinking; to Malcolm Haddon who always had time for me, for helping me to explore and organise my thoughts, and for very promptly delivered and helpful comments on my drafts; to Andre Punt without whose encouragement I would not have begun this PhD, for underpinning genius, and for giving up his Thanksgiving holiday to give me valuable feedback on my draft; to Keith Sainsbury who helped me start off the project; and to my earlier supervisor Doug Butterworth for starting my career. Barry Bruce, of the Commonwealth Scientific and Industrial Organisation (CSIRO), triggered my interest in the topic of modelling movement behaviour, provided me with the satellite tracks used in this thesis, and gave me many useful ideas during the course of several stimulating discussions. My friends have sustained me, and have put up with my absences - to Mitchell Rolls, who has enough research, administration and students of his own to keep him very busy but who nevertheless found time to help me organise my despairing thoughts into a sustaining schedule; to my Australian mum, Robin Burns for a strong shoulder, and wise advice including the reassurance that nobody makes noticeable progress during the first year of their Phd; and to my friends and colleagues at the CSIRO for both technical and practical help, in particular, Jeff Dunn for his help in accessing GeoSciences Australia bathymetry data that were stored in a CSIRO database. I cannot find adequate words to thank my parents for their lifelong support, patience, wisdom, listening ears and for, incredibly, being happy to proof-read my final draft. Finally, my thanks go to the University of Tasmania, the CSIRO and the Quantitative Marine Science program and its head, Richard Coleman, for financial, administrative and practical support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

GENERAL INTRODUCTION

Understanding animal movements and identifying the factors that influence these movements has long been a goal of ecological research (Turchin, 1998; Okubo, 1980; Okubo & Levin, 2001). Better understanding of animal movements can improve natural resource management through mitigation of unwanted bycatch (Broekhuizen *et al.* , 2003), guiding the selection of new protected areas (Maury & Gascuel, 1999; van Vuren, 1998), and forecasting the impact of climate change on population distributions (Bowler & Benton, 2005).

Information on the movements of animals can be obtained from a number of sources, even examination of their tissues. Otolith microchemistry, for example, may reveal which bodies of water a fish has resided in during its life, by the isotopes found in sections of its ear bones (Elsdon & Gillanders, 2003). Mitochondrial and nuclear DNA investigation revealed that while male white shark DNA is shared between Australia and Africa, female genetic material is not, indicating that males travel across the Indian ocean whereas females are resident (Pardini *et al.* , 2001). However, the most powerful techniques for revealing animal movements are tagging and tracking methods. A single track of a female white shark showed that she travelled from South Africa to Australia, and subsequent photo-identification that she returned to South Africa. Thus the conclusions of the genetic study were overturned: although only male genetic material is exchanged it is the females that travel to collect it and carry it back home (Bonfil *et al.* , 2005).

Most statistical methods require large numbers of independent observations from which inferences may be drawn. A single track, composed of non-independent observations taken from a single individual, is unsuitable for application of most statistical methods other than those of time series methodology (Chatfield, 2004). Here the

track itself, a sequence of serially correlated observations, is the time series.

Further difficulties must be overcome: time series methods have been developed primarily for identifying patterns evident within a long series of sequential observations, so that this pattern can be forecast into the future. The underlying reasons for this pattern, and its covariates, are often of secondary importance. An annual seasonal oscillation, for example, may be evident and can be projected into future years. In order to detect such a seasonal pattern, observations will be required from at least two, usually many more, past years. Seasonal patterns are very likely to occur in animal movement behaviour, but animal tracking data are usually measured in weeks or months, not years. Such tracks are therefore unlikely to yield repeated observations of past patterns, but rather a single instance, or perhaps only a part of, a particular pattern. Tracks must therefore be understood, not by recognising a repeating pattern, but rather by identifying the external covariates or internal motivations that drive the moving animal, so that future movements may be forecast using future projections of these covariates (such that those forecast by oceanographic or meteorological models). Time series techniques typically do not make use of large datasets of covariate information because such correlates of the repeating patterns found in the data are of secondary importance. In the case of animal tracking data, however, these covariates are of importance.

A family of time series methods is presented in this thesis that is capable of combining covariate information with observed tracks. Movement behaviour, as a response to measured environmental variables or to theorized internal states, is hypothesized by the investigator. Alternative behavioural hypotheses are formulated, and the support given to each by an observed track is measured. As further tracks are collected, and hypotheses tested, theories of movement behaviour may be refined. In this way, it is hoped, understanding of the, no doubt, complex movement behaviour of animals will be gained incrementally.

Databases of satellite tracks are growing around the world (Coyne & Godley, 2005) and statistical methods capable of using these tracks for inference about movement behaviour are badly needed. It is not expected that any theory of animal behaviour, implemented as a statistical model, will be exactly correct. Instead, it is hoped that the process of model selection and model building will facilitate incremental learning regarding the movement behaviour of animals. As Box (1979) famously said 'all models are wrong but some are useful'. This thesis presents a statistical model for animal movement behaviour that can be used together with methods of model comparison and observed animal tracks to make inference about the motivations of

movement behaviour. Hypothesized movement behaviours are unlikely to exactly replicate the motivations of a moving animal, instead the aim of this work is to provide a tool that measures the support given by track data to hypotheses of movement behaviour so that by discarding and refining hypotheses a greater understanding of movement behaviour can be gained incrementally.

This chapter briefly discusses the four elements required for this process: track data; covariate information; a model capable of using a hypothesis of movement behaviour, together with candidate covariate information, to simulate the track of a moving animal; and a means of measuring the support given by an observed track to the model generated simulation. Finally, an outline of the thesis is given.

1.1 Track data

Track data have been collected by direct observation (by eye or using a camera) or through radio- (or sonar-) tracking, either by an investigator physically following a tagged animal, or through arrays of monitoring devices placed in the region in which an animal moves. More recently, however, the tracking of moving animals, even those that travel great distances across oceans or that fly over mountain ranges, has been made easier by advances in satellite tracking technology (Priede & French, 1991; White & Garrott, 2006). Researchers no longer have to follow along after the animals they are tracking, holding radio-receivers. Satellites and GPS systems provide accurate, frequent position fixes for animals that move on the surface of the land or ocean. Satellite communication is not possible for those that move beneath water, seldom or never surfacing. For these animals, archival tags are required. These store information, including light-level data from which location may be inferred, and must either be retrieved or after a pre-determined time, detach themselves from their host animal, float to the surface, and send their information to satellites. This technology, although expensive, is now widely used. Coyne & Godley (2005) report that between 1995 and 2005 a five-fold increase was reported in the numbers of animals tracked by the Argos satellite system. However, due to the expense of this technology, available tracks typically represent only a small number of individuals from any population.

A track, as defined in this thesis, is a sequence of location estimates for a moving animal that is serially correlated and is often recorded at irregular intervals. Tracks are serially (auto-) correlated because the location of an animal, an hour

hence, has more to do with its present location and its speed of travel than its ultimate destination. Satellite tracked locations are not equally-spaced in time because the tags communicate with satellites when possible, they cannot do so when the animal to which they are attached is submerged, or inside a rocky cave, or when no satellites are overhead (Priede & French, 1991). Radio-tracking is also likely to provide irregularly-timed location estimates and even archival tags, which record information at regular intervals, are only able to provide information from which location can be estimated at dawn and dusk, provided the animal is not diving too deeply and cloud cover is not too great, so that even these tags may provide somewhat irregularly-timed locations.

The error structure of observed track locations can be somewhat unusual. The Argos system grades every location that it provides into one of six error classes. The error structure associated with each class has been shown to be normally distributed but ellipsoidal, having greater variance in longitude than in latitude; the extent of this variance differs among error classes (Vincent *et al.* , 2002). Tracks derived from archival tag light level information also show greater errors in longitude than in latitude (Sibert *et al.* , 2003).

1.2 Covariates and hypotheses of behaviour

The satellite technology used to collect track data has also facilitated the collection of environmental information. The MODIS system, for example, measures sea surface temperature and colour for the oceans. Measurements of ocean height are used to infer movements of ocean currents. These data have been used to develop accurate, mid-scale ocean and atmosphere models as well as maps of terrestrial conditions (Campbell, 2006; Ryerson, 1998). Such environmental information is typically available in the form of 2- or 3-dimensional grids.

Moving animals may respond to particular environmental cues, for example, tuna appear to move so as to remain in waters of particular temperatures (Block *et al.* , 1998). Observed tracks of moving animals may contain cues that would allow investigators to infer which environmental variables are possible covariates of movement behaviour, and which are not.

One way of inferring whether an environmental variable is a determinant, or at least a correlate, of movement behaviour, is to overlay observed tracks on a map of

environmental information and to see, by eye, whether there seems to be any correspondence. Most investigations will at least begin with an informal investigation of this sort, but scientific enquiry requires more rigorous, formal statistical testing of hypotheses. Therefore, a statistical method is required that can relate two very different sorts of information: tracks, and maps of environmental covariate information.

Of course, moving animals will not only respond to observed and observable features of their environment, they will also respond to features that investigators have not observed, such as the presence of prey animals or conspecifics, and they will respond to internal motivations such as a desire to move to a spawning area during a breeding season. Some of these behaviours might be included, implicitly, in a model of movement behaviour, as random noise. More persistent behaviours, such as migration, will also have to be accounted for. The statistical model for movement behaviour would therefore also need to incorporate behaviours that do not relate to observed covariate information, such as migrations or movements towards spawning grounds.

It is envisioned that investigators wishing to use track data to make inferences about the motivations for movement behaviour will form competing hypotheses. Some of these will relate to covariate information, others will identify areas towards, or away from, which the animal is thought to move at particular times. The same covariate information may be used, in different ways, by competing hypotheses that predict different responses. For example, it may be thought that a particular animal uses the earth's magnetic field as a navigational aid. It may do this naively, moving 'upstream' through the magnetic field until it finds an area that matches the magnetic properties of the location towards which it wishes to move, or it may have a memory of the irregularities in the field between its starting point and target location and may correct for these. It is also recognized that different hypotheses regarding the motivations of a moving animal may lead, indistinguishably, to similar theorized behaviour. For example, a marine animal moving along the east coast of Australia could be following a line of bathymetry, could be using the sounds of coastal waves to remain a particular distance from the coast, or could be maintaining an orientation with respect to the earth's main magnetic field. These factors appear to be almost perfectly correlated in this region, so that these three hypotheses are unlikely to be distinguishable using track data.

More than one of the behaviours discussed above could occur simultaneously - an aquatic animal may undertake a migration, navigating using the earth's magnetic field, and may respond to preferred water temperatures encountered on the way by

slowing down, perhaps to feed.

1.3 Modelling framework

Combined with present computing power, remotely sensed information gives researchers the opportunity to investigate which of the measured variables of the environment through which an animal moves, appear to affect its movement behaviour.

A numerical model capable of combining track data with hypotheses regarding behaviour (which may use covariate information) must couple very different types of information: tracks, maps of covariate data, and pure theory. In addition, the model must be robust to responses to unobserved factors for which no specific theory is available, such as responses to other, untagged, animals. Movement behaviour is likely to be highly complex and individuals are unlikely to all behave in the same manner. Theoretical models can therefore be expected to be significantly in error, so that process errors should be modelled. Observation errors should also be modelled; track locations, even when collected using GPS enabled satellite tags, are imprecise. These observation errors must be unequal in latitude and longitude. Such a model must also be able to account for the irregular time steps typical of non-archival tracking data. In addition, it would be useful to have the facility to estimate latent (unobserved) locations. Long delays can occur between subsequent observed locations and, because the assumption is usually made that animals move in a straight line between observed locations, this can sometimes result in unrealistic behaviour. For example, an aquatic animal may return a location estimate from one side of an island or spit, and its next location from the opposite side, so that a straight-line interpolation would indicate that it had crossed the land barrier. Even without an intervening barrier, the hypothesis of movement behaviour implemented by the model might predict a more complicated path between two distant points than that given by a straight line. For these reasons, it would be desirable to have the ability to include latent locations in a model of movement behaviour.

State space models (SSMs) provide a promising framework for overcoming these difficulties (Patterson *et al.* , 2007; Hilborn, 1990; Newman, 1998, 2000). SSMs represent a system as a time series of unobserved states (such as the true location of a moving animal), and a separate, but related, set of observations of these states. This structure allows estimation of errors in both the observations, and in the process that gives rise to the unobserved states (process errors) (Harvey, 1990). The SSM

will model the track as a time series, therefore taking account of its inherent serial correlation. Each track would be modelled individually, therefore, data from a large number of individuals are not required. However, tracks from several individuals can be combined in a hierarchical framework (Jonsen *et al.* , 2003; Newman, 2000; Newman & Lindley, 2006). Sibert *et al.* (2003) and Jonsen *et al.* (2006) have shown that the unusual error structure of satellite tracking data can be accommodated within an SSM framework.

Some challenges remain. SSMs represent time in a discrete fashion, giving the state of a system at a series of equally spaced points in time. Satellite tracking provides data that are irregular as it relies on the animal surfacing and the tracker signalling an overhead satellite. Others have applied SSMs to satellite tracking data by 'regularising' the data through calculation of some form of weighted average (Flemming *et al.* , 2006), by assuming that all observations occurring within a time step were taken at the same time (Jonsen *et al.* , 2006), or by interpolation (Tremblay *et al.* , 2006). Jonsen *et al.* (2005) represent irregular time steps in the model itself through the observation equation. Their observations of location are time-weighted averages of an underlying regularly-timed series of unobserved locations (the state variables). As an alternative, this thesis presents an SSM framework in which locations are irregularly spaced in time, and the duration of each time step is accounted for within the model's state equation.

Published movement models applied to tracking data typically include only a single potential covariate for movement, if any. Jonsen *et al.* (2003) made the variance of the process errors a function of sea surface temperature (SST) so that turtles slowed when they entered warmer water. Behaviour, not mediated by a covariate, has been included as a tendency to drift in a particular direction (Jonsen *et al.* , 2005, 2006), as switching between alternative parameter sets (Jonsen *et al.* , 2005; Morales *et al.* , 2004), or as deviation from a pre-specified route (Flemming *et al.* , 2006). However, behaviour is likely to result from a combination of all of these factors operating at once, and perhaps a need to follow a navigational route as defined, for example, by landmarks, the earth's magnetic field or star patterns.

This thesis presents a method of representing movement behaviour as one or more advection fields that push the moving animal in particular directions. Separate models (co-models) use observed covariate information and theories on how and when animals choose to move, to generate 2-dimensional sets of advection forces. These advections represent the direction in which the theory, possibly coupled with observations of the environment, indicates that the animal will move. Any num-

ber of advection fields can be incorporated into the model - the overall direction of movement is given by a weighted, linear combination of advection forces.

The influence of the environment on the movements of the animal causes future locations to be dependent on past locations (and the state of the environment at the past locations) so that the problem becomes highly nonlinear. In the past, the usefulness of SSMs has been limited by difficulties in inference for nonlinear, and non-Gaussian, forms of this model. Now Markov chain Monte Carlo (MCMC) methodology and current computing power offer release from these constraints (Buckland *et al.* , 2004; Newman & Lindley, 2006; Thomas *et al.* , 2005; Jonsen *et al.* , 2003, 2005, 2006) although some challenges remain and this is an area of active research (Buckland *et al.* , 2004; Newman & Lindley, 2006). A great advance has been provided by the WinBUGs software (Spiegelhalter *et al.* , 2004), which allows users to specify even quite complicated models, making inference about their parameters using MCMC methodology. This software has been used to implement SSMs in population dynamics (Meyer & Millar, 1999b) and movement contexts (Rivot *et al.* , 2004; Jonsen *et al.* , 2006; Morales *et al.* , 2004; Thompson *et al.* , 2005). However, this software, currently, does not allow the incorporation of potentially large datasets of covariate information, or the additional coding required to represent complex movement behaviours. Buckland *et al.* (2004), Newman & Lindley (2006), and Thomas *et al.* (2005), in companion papers, describe a method of inference for SSMs using Sequential Importance Sampling (Liu & Chen, 1998; Liu, 2004). However, a serious drawback of this method is 'particle depletion', which leads to overestimation of high density regions of the posterior and underestimation of low density regions. One of the authors' proposed solutions to the problem is to use MCMC. This thesis presents an SSM model implemented in the R statistical software (R Development Core Team, 2007) that uses MCMC methodology for statistical inference.

Moving animals have been observed to show inertia, a tendency for consistency of direction (Roberts *et al.* , 2004; Guilford *et al.* , 2004; Bartumeus *et al.* , 2005), and this too must be incorporated into realistic models of movement. Inertia can be imposed on the direction of motion by adding an additional layer to the state equation, making it a structural time series.

Movement models typically assume that animals travel in straight lines between observations (e.g. Jonsen *et al.* , 2006; Kareiva & Shigesada, 1983; Sibert *et al.* , 2003). This assumption can cause difficulties, such as when an obstacle (such as a landmass in the case of an aquatic animal) intervenes and when, as often happens for aquatic animals that do not need to surface to breathe, long delays occur

between observations of location. The model presented in this thesis is capable of estimating latent locations so that a probabilistic estimation of actual location, even at unobserved times, is obtained.

Behaviour may change seasonally, or with maturity. In this thesis, this is modelled by allowing the coefficients for the advection fields to be functions of time. In addition, the changes in these coefficients can be correlated, so that as one grows in importance, another diminishes. For example, the end of a breeding season can be represented by the combination of an advection field for movement away from a breeding ground, with another for movement towards it. A negative correlation between these coefficients will allow movement away from the breeding ground to strengthen over time while that for movement towards it weakens. Thus, by the end of the breeding season, the net direction of movement would have reversed.

A common problem of movement models that do not explicitly model navigation, is a tendency for simulated animals to become trapped in semi-enclosed parts of the landscape and, conversely, to step across barriers narrower than their step length. If navigation is under investigation, then the advection field that is generated by a co-model implementing a navigational method, might steer the moving animal around obstacles. For example, an animal that navigates by following a line of bathymetry will not approach shallow bays in which it might become trapped, and would not approach narrow peninsulas closely enough to step across them. However, if the navigational method does not incorporate a means of avoiding obstacles, or if navigation is not explicitly of interest, then some other means is required to avoid this problem. This commonly encountered problem has not been given much attention in the literature. This is presumably because studies of movement behaviour that are concerned with where animals choose to move and which covariates trigger their movements, are not focussed on navigation - how animals get where they are going. Entrapment may appear to be a modelling artefact. This thesis draws attention to the importance of this problem and suggests a method whereby Laplace's equation for diffusion may be used to generate an advection field that flows, and guides the moving animal, around obstacles. This does not entirely solve the problem of stepping across narrow barriers but it is hoped that this work will at least encourage explicit recognition of the problem and encourage debate that may lead to other solutions.

1.4 Hypothesis selection

The fourth, and final, aspect necessary for inference about movement behaviour, is a means of measuring the support given by an observed track to a model that represents a hypothesis concerning movement behaviour. It is unlikely that any model of movement behaviour will closely resemble the real, complex, behaviour of the tracked animal - in particular, responses to unobserved covariates cannot be modelled, some behaviours may be very complex, and some cues may be very local. Therefore, point estimates of the parameter values of the movement model may not always be of great interest. Bayesian methods, which provide marginal probability distributions for parameter values, seem more appropriate. It is expected that the model will be used to discriminate between broadly different hypotheses of behaviour. The Bayes factor is used to provide a measure of support given by the track data to alternative hypotheses of behaviour, each represented by a movement SSM that uses a particular set of advection fields that were generated by co-models of theorized behaviour.

1.5 Thesis structure

Chapter 2 reviews the literature on methods of collection of track data, and methods of quantitative analysis of movement with emphasis on state space methods, and places the work presented in this thesis within that context.

Chapter 3 describes an SSM framework for modelling a track whose sequence of location observations may be irregularly-timed. The model framework includes both observation and process errors, and allows for inertia and for unobserved (latent) locations. Movement behaviour is represented by advection fields that are calculated by co-models, external to the main movement model. This allows great flexibility and for the representation of a wide range of behaviour types as well as the use of datasets of potential covariates for movement. Examples are given of the ways in which behaviour might be represented as advection fields, and simulation tests are presented to illustrate the properties of the model. The MCMC method used to estimate the posterior distribution of the model is described.

Chapter 4 extends the state equation of the SSM model to allow behaviour to change with time. Simulations are presented that explore the accuracy of param-

eter inference in scenarios in which one behaviour diminishes in importance over time while another, correspondingly, increases in importance.

Chapter 5 describes the Bayes factor and several methods for approximating the marginal likelihood, a difficult integral which is required for calculating the Bayes factor. This chapter shows that, in simulation, model comparison measures can select the movement behaviour that was used to generate the data, over a similar candidate behaviour. The deviance information criterion (DIC) is a more robust measure, in this regard, than the Bayes factor, which is sensitive to the choice of prior.

Chapter 6 shows that the harmonic mean estimator of the marginal likelihood (the normalising constant for the posterior) provides estimates that are insufficiently accurate for practical model comparison. This finding is in contrast with advice given in the literature. A manuscript based on this chapter has been accepted for publication by the *Australian and New Zealand Journal of Statistics*.

Chapter 7 demonstrates that Laplace's equation can be used to generate an advection field that will guide moving individuals around obstacles and semi-enclosed areas such as bays. Use of the heat equation presents a solution to, and highlights a common problem experienced by, those modelling moving individuals.

Chapter 8 validates the modelling techniques presented in thesis by an application of the movement modelling and model comparison framework to observed tracks of white sharks in South Australian waters. It was shown that the hypothesis that white sharks navigate by following a line of bathymetry is not supported by the data.

A general discussion of the work presented in this thesis appears in the final chapter, Chapter 9, where avenues for future work are discussed.

Appendix 10 lists the symbols used in Chapters 3 and 4, together with a short description, and their dimensionality.

The MCMC method (a variant of Gibbs sampling) used in this thesis requires the calculation of conditional distributions. Appendix 11 describes the general methodology used to calculate these from the posterior distribution, and subsequent appendices present specific calculations for a gamma conditional (Appendix 12), a multivariate normal conditional (Appendix 13), and a Wishart conditional (Appendix 14) distribution. All of these calculations pertain to the conditional dis-

tributions used in Chapter 4, however those used in Chapter 3 were derived using the same methodology. Each appendix presents the most complex example, for the relevant distribution, of the conditional calculations performed as part of this work. The other conditionals presented in Chapters 3 and 4 can therefore be regarded as subsets of the calculations presented.

The simulation presented in Chapter 6, to illustrate that the harmonic mean estimator of the marginal likelihood is inadequate for practical model comparison, is an extremely simple one that can be implemented in the R programming language (R Development Core Team, 2007) in a few lines. This code is given in Appendix 15.

CHAPTER 2

Literature review: Quantitative analysis of track data with emphasis on state space models

2.1 Abstract

This chapter provides definitions for terms used throughout the thesis, and gives an overview of relevant parts of the literature. Track data have been collected by direct observation, through use of acoustic and radio tags, using archival tags and, most recently, using GPS-enabled tags that transmit position information to satellites. Datasets of track information are accumulating around the world and empirical methods are needed that can be used, together with track data, to make inference about animal movement behaviour.

The use of track data presents a number of challenges. The sequence of locations that constitute a track is highly autocorrelated. Both observation and process error are likely to be substantial and should be modelled separately because information is often available on the structure of observation error. Location fixes occur when the investigator is close to the animal (for radio-tracking) or when the tag is able to communicate with overhead satellites (for GPS tags) so that time steps are typically unequal. Large gaps can occur between location estimates, particularly in the case of marine animals which may spend long periods submerged.

A model capable of making inference about movement behaviour would need to be flexible enough to model a wide variety of theorized, possibly complex behaviours.

It is necessary to calculate some measure of the support given by the data to competing theories of animal movement behaviour.

Quantitative methods applied to track data and to tag-recapture data include a variety of empirical (mostly descriptive) methods, random walk and advection-diffusion models, bulk transfer processes and individual-based models. A promising area for modelling the highly autocorrelated tracks is time series modelling and, in particular, state space models (SSMs).

SSMs account for the autocorrelation in the track, they model observation and process error separately, and their ability to use structural time series gives the model builder great flexibility in the behaviours that can be incorporated into the model. Historically, the use of SSMs has been limited by difficulties in inference. Great advances in numerical integration using the powerful Monte Carlo methods have offered a solution to this problem, encouraging the use of SSMs in a wide range of fields. Examples are given from the field of wildlife population modelling and animal movement modelling. Solutions for challenges regarding inference for SSMs are still being found and this thesis attempts to contribute to that field. A Bayesian framework facilitates the estimation of a posterior probability distribution for the location of the animal throughout the tracking period, rather than yielding a single, optimal, track.

The preferred method of model comparison for Bayesian models is the Bayes factor, which involves a difficult integral that can be estimated using a number of techniques. Simpler, but possibly less accurate methods are available but of these, the most suitable for application to SSMs is the deviance information criterion DIC.

2.2 Introduction

Observations of animal movements can be divided into spotting and tracking. *Spotting* involves noting the presence of animals at particular locations at particular times without following them as they move. For example, it might be noted that swallows are present in the southern hemisphere only during the austral summer and in the northern hemisphere only during the boreal summer. The resulting hypothesis that swallows migrate between hemispheres at the end of summer could be tested by a more sophisticated method of spotting: marking individual birds and recapturing them, repeatedly, in different hemispheres. Such mark-recapture techniques depend on an ability to recognize individual animals or groups, either

through existing natural markings such as the pattern of scars on the skins of sharks (Bonfil *et al.* , 2005) and whales (Graham & Roberts, 2007), or through application of artificial marks such as tattooing, branding, clipping, and tagging (Turchin, 1998).

Spotting methods are useful for inferring movements such as migration and dispersal, where an individual remains in one location for some time, and then moves, relatively rapidly, to another location where it again resides. However, spotting methods are unlikely to provide insight into day-to-day movements, such as foraging or perhaps mating behaviours. These are likely to result from a mixture of internal motivations, such as hunger and a desire to return to a feeding area that has yielded success in the past, and responses to external factors such as stumbling across a favourable feeding area, or preferring particular habitat types, a motivation attributed to tuna (Block *et al.* , 1998). Cues for such day-to-day movements may instead be inferred from the results of *tracking* techniques. Here, animals are followed as they move, either physically or electronically, and their location is measured with sufficient frequency to yield a track, which is defined in this thesis as an autocorrelated set of sequential timed locations.

A track might consist of movements occurring at three scales. First, small movements, or darts, will occur on a scale similar to that of the track's observation errors. Although these will doubtless have some cause, even if it is a purely neurological one (Turchin, 1998), the presence of observation errors will render such small-scale movements prohibitively difficult to study. Second, the track will reveal medium-scale day-to-day movements such as foraging. Third, the track may capture part of a large-scale movement, such as a migration, that occurs at a scale larger than the track itself. Tracks therefore lend themselves best to the study of the medium-scale, day-to-day movements that occur at a scale greater than that of observation errors but smaller than the scale of the track itself. Nevertheless, any highly directed large-scale movement that might be occurring during the tracking period would have to be explicitly accounted for in a model of movement behaviour, probably as a preference for movement in a particular direction. From the point of view of an investigator attempting to uncover the causes of medium-scale, day-to-day movements the smaller-scale movements can be regarded as random noise.

Spotting methods have been available for a long time so that migratory movements are relatively well understood and mathematical methods for simulating them have had time to develop (e.g. Sibert & Fournier, 1994; Sibert *et al.* , 1996, 1999). Tracking by direct observation has long been available for insects because the space in which they move is sufficiently small to allow observation by a stationary camera,

or to allow a human to follow without risk of being left behind. Insect movements seem to be well described by correlated random walks (Kareiva & Shigesada, 1983) or by purely random walks and therefore by their continuous analogue, the diffusion model (Turchin, 1998).

Tracking of larger, and therefore more widely ranging and faster moving, animals is dependent on more sophisticated technology. Radio and ultrasound equipment has allowed investigators to follow tagged animals, either on foot or from a vehicle or vessel, but the need for the presence of an investigator may alter the behaviour of the animal and limits the duration of tracking. Since the 1970s, tags have been available that obtain and transmit accurate position data using satellites so that tagged animals can be tracked remotely for a period of time that is limited only by battery life. Satellite remote sensing also yields datasets of variables such as sea surface temperature and terrestrial vegetation type that may be determinants of movement - at least those movements occurring at the medium-scale. In addition to providing covariate information for inference using tracks observed in the past, remotely sensed information has been used in the development of three-dimensional computer models of ocean and weather conditions from which forecasts are available for these potential movement determinants (e.g. Oke *et al.* , 2005; Brassington *et al.* , 2007). With all of this information, better understanding and prediction of the movements of animals should be possible. However, much of this data is relatively new and the techniques required for coupling the highly autocorrelated, individualistic, tracks of single animals with maps of environmental conditions and theories about how animals might be behaving, are still in early stages of development.

The purpose of this chapter is to review animal tracking methodologies and published quantitative analyses of tracking data. The chapter begins by defining some of the terms that will be used throughout the thesis, then outlines tracking technologies, and finally discusses the types of quantitative analysis that have been applied to tracking data. The argument is made that state space models (SSMs) offer particular promise for modelling movement behaviour (Jonsen *et al.* , 2005; Patterson *et al.* , 2007), and this framework is therefore given greatest attention in the review. A brief outline is provided of the rather confusing nomenclature of SSMs, along with details of the SSM framework, and a discussion of practical application of SSMs, which includes their troublesome inference, published applications of SSMs to movement modelling, and hierarchical SSMs. Finally, a discussion is presented on methods of model comparison with emphasis on those applicable to SSMs. This work has an aquatic and, in particular, a marine focus but the computations presented in this thesis are equally applicable to the movements of terrestrial animals.

Terrestrial references are also given, where possible.

2.2.1 Definitions

The term *track* is used in this thesis to mean a sequence of observed positions of a moving individual, with these observations sufficiently close in time for the sequence to be autocorrelated. *Path* indicates the true positions of the moving individual. The (true) path differs from the (observed) track due to observation errors. Tag-recapture data are not considered to constitute a track because recaptures are typically made sufficiently far apart in time for locations to be uncorrelated. These data tend, rather, to be correlated with the location of the person or instrument that observes the presence of the animal.

The term *moving* is used in this thesis to indicate all forms of travel including migration, dispersal, foraging and small seemingly random movements. Unlike the work of Turchin (1998), this thesis does not distinguish between forms of movement because the modelling framework presented in later chapters encompasses movement at all scales.

The term *movement behaviour* is used to indicate changes of location by an individual in response to external and internal cues. External cues are aspects of an individual's environment such as temperature, habitat type and ocean currents. Internal cues are motivations such as a desire to move towards a known feeding or pupping ground. Social behaviours, such as attraction to, or avoidance of, other individuals is not something explicitly considered in this work. The modelling framework presented here has been developed for application to a single observed track; however, it could be extended to a hierarchical form so that several tracks could be considered. In this form, interactions between tagged animals might conceivably be examined. The response of a tagged animal to an another, untagged, animal, cannot be inferred without supplementary information regarding the presence of the untagged animal.

The term *covariate* is used in this thesis to indicate any measurable (usually environmental) quantity that may be a determinant of movement behaviour or may, at least, show correlation with movement behaviour. It is understood that such correlation may not be indicative of a causal relationship.

A *random walk* is a simple model that describes movement as a series of steps, each drawn from a probability distribution. The successive steps of a *purely ran-*

dom walk are independent, whereas those of a *correlated random walk* are correlated with one another (Turchin, 1998). For example, Brownian motion is a particular form of purely random walk in which each step is drawn from a normal distribution (Chakravarti, 2004). A Brownian motion in 2-dimensional space uses a bivariate normal distribution. When step lengths are drawn from a Levy distribution the result is another type of purely random walk called a Levy flight (Klafter *et al.* , 1996; Shlesinger, 2001; Chakravarti, 2004). When the Levy distribution is used to describe the velocity of steps, rather than their length, the result is called a Levy walk. Correlated random walks occur when the distance or direction of the previous step affect those of the next. Animals often display directional persistence, or inertia, so that purely random walks may not be appropriate descriptors of animal movement (Bartumeus *et al.* , 2005).

Objects following any of these forms of random walk are described as *walkers*.

2.3 Collecting track data

Track data may be collected by a variety of methods.

2.3.1 Direct observation

Quantitative analyses typically require large volumes of data; something that has not historically been available in the field of animal movement behaviour because, prior to the invention of satellite tracking devices, this data typically had to be collected by the labour intensive means of following the animal. An exception to this are observations on insects (Turchin, 1998) whose movement behaviour has been relatively well studied, presumably because they are relatively easy to follow. Many insects are large enough to see while small enough to move without severe restriction within the confines of a room. Those studied within a laboratory setting, where the area in which they can move is bounded, can be followed using a camera (Jeanson *et al.* , 2003; Wiktorsson *et al.* , 2004). Many of the more mobile insects, such as butterflies, are nevertheless slow enough to be followed on foot (Root & Kareiva, 1984). Insect populations are typically numerous, therefore lending themselves to mark-recapture studies, which, while not yielding track information, can be used to infer large-scale migrations.

2.3.2 Aquatic acoustic listening stations

Passive acoustic listening stations are placed, stationary, in aquatic areas, often in arrays. They are capable of detecting the passage of appropriately tagged animals within a range of 600 to 800m (Heupel *et al.* , 2006). Although it is possible to use arrays of at least two listening stations to assess the location of a tagged animal by triangulation and therefore to track it, such tracks are typically small-scale due to the range limitation of this method. A large number of listening stations would be required to broaden the scale. Nevertheless, this method can be of use for smaller animals such as rock lobster (Frusher *et al.* , 2003). A similar method, which has a longer detection range, has been applied off the coast of Alaska where an array of hydrophones records the vocalizations of migrating bowhead whales (Clark & Ellison, 2000).

2.3.3 Radio and acoustic tracking

Radio-tracking devices allow investigators to use hand-held, or vehicle mounted receivers to follow animals to which they have attached a radio transmitter (White & Garrott, 2006; Kenward, 2001; Mech & Barber, 2002). Acoustic (ultrasound) transmitters have been used for following marine animals (Sundstrom *et al.* , 2001; Strong *et al.* , 1992, 1996; Goldman & Anderson, 1999) but for brevity the term 'radio' is used, here, inclusively. By this means, free-ranging animals can be tracked provided they are large enough to carry the device, and provided the investigator is able to follow. Radio devices have been attached to birds as small as grouse (Boag, 1972), and herring gulls (Amlaner *et al.* , 1979), and to mammals as small as lemmings (Brooks & Banks, 1971) and voles (Leuze, 1980). For larger animals, the limitation is not the animal's ability to carry the transmitter but rather the investigator's ability to follow. Strong *et al.* (1992, 1996) required a small vessel to follow nine ultrasonically tagged white sharks for periods of up to 28 hours. Animals that undertake large migrations can only be followed for a small part of their route and flying animals are particularly difficult to follow.

Tagged animals can be tracked by triangulation, from two stationary elevated platforms (White & Garrott, 2006). While this does away with the need to physically follow the animal, it limits the tracking area to the range of the radio receiver.

Quantitative analyses of radio-tracking data have concentrated on methods for home range estimation (Voight & Tinline, 1980; Larkin & Halkin, 1994; Lawson

& Rodgers, 1997), identification of preferred habitat, and estimation of population size and survival rates (White & Garrott, 2006; Becker, 1991).

Radio-tracking allows tracking of animals that are more wide-ranging than insects but, like direct observation, is labour intensive. It is limited to animals that can be physically followed, or that do not wander out of range. Models of movement behaviour that make use of track data will typically assume that the track is representative of natural movements, unaffected by observation. This is unlikely to be true of radio-tracked animals that have been physically followed by investigators who are close enough to detect the signals from their tags.

2.3.4 Archival tags and Satellite trackers

Archival tags are microprocessors, attached to animals, that regularly record and time stamp certain features of the environment such as light intensity and water temperature (Horning & Hill, 2005). Less sophisticated archival tags are not satellite enabled, they have to be retrieved before their stored information can be downloaded. However, newer archival tags detach themselves after a pre-programmed period of time and float to the surface of the ocean where they transmit their stored information via satellite (Block *et al.* , 1998). Those that self-detach and transmit to satellite are called pop-off Archival Transponding Tags (PAT). Daily location can be inferred from the light level data by noting the timing of sunrise and sunset. These tags are deployed on aquatic animals that seldom, or never, surface because satellite links cannot be established during underwater transit.

Terrestrial animals, and aquatic animals that frequently surface, permit the use of attached devices that frequently transmit radio signals carrying location information to overhead satellite networks, such as Argos (Priede & French, 1991). Satellite tracking of animals using these Platform Terminal Transmitters (PTTs), or GPS (Global Positioning System) tags, was first reported in the early 1970s (Beuchner *et al.* , 1971). Initially bulky, the earliest satellite trackers could only be deployed on large animals such as bears (Craighead *et al.* , 1971; Lentfer & DeMaster, 1982). By the mid-1980s trackers had become smaller and the first seabird study became possible (Parmalee *et al.* , 1985). Now GPS trackers can be attached to animals at least as small as pigeons (Guilford *et al.* , 2004). The duration of the tracking period is limited only by the battery life and durability of the transmitter; investigators can follow animals from their office computer, and the behaviour of tracked animals is influenced only by the trauma of capture and the presence of the transmitter,

not by the presence of a radio-receiver bearing human following along behind them. Location is estimated using GPS, usually with greater accuracy than that inferred from the light data collected by archival tags. Satellite tracking is a much less labour intensive, and less invasive, method of tracking animals than its predecessors. While the presence of the tag, or the trauma of capture and release, may alter the behaviour of the tracked animal, this problem is inherent in all other tagging methods except for the non-invasive practice of recognising individuals using natural markings, and even then an observer is usually present. Satellite tags are expensive, however, and their batteries typically last for only weeks or months so that conventional tagging is still preferable for investigations that do not require accurate, frequent, position fixes.

Datasets of tracking information for medium and larger terrestrial and marine animals are accumulating around the world. Between 1995 and 2005 a five-fold increase was recorded in the number of animals tracked by the Argos system (Coyne & Godley, 2005). These animals are unlikely to move in the instinctive, programmed, way typical of insects so that the models developed for insects are likely to be inappropriate for application to satellite-tracked individuals.

Not only have satellites advanced our animal tracking ability, they have also allowed the measurement of environmental variables (to which moving animals may be responding) on a global scale. For example, the MODIS system (Shutler *et al.* , 2005) measures sea-surface temperature and colour (a proxy for phytoplankton and hence productivity) (Campbell, 2006; Ryerson, 1998) for the oceans as well as monitoring the land. These data have not previously been available so that new methods must be developed for incorporating such map-based data into ecological models.

The sophistication of data collection technology has outpaced that of analytical methodology - new methods are needed.

2.4 Quantitative methods

2.4.1 Empirical methods

Empirical approaches to movement analysis are those that yield generalizations about movement behaviour through calculation of summary statistics from track

data (Austin *et al.* , 2004). For example, track data have been used to calculate the statistical properties of individual steps (e.g. Coughlin *et al.* , 1992; Lowry *et al.* , 2001) - properties such as the distances moved (step lengths) and directions or turning angles between successive observed locations.

Empirical studies are often focused on understanding search strategy by examining the efficacy of different kinds of walk for finding scattered targets such as food plants (Bartumeus *et al.* , 2005). Zollner & Lima (1999) found that correlated random walks are better than purely random walks as search strategies for finding habitat patches in a fragmented landscape. Levy flights have been shown to provide an optimal search strategy for locating sparsely distributed targets (Viswanathan *et al.* , 1999; Bartumeus *et al.* , 2005). Bertrand & Atiquipa (2005) showed that the movements of anchovy fishers, searching for schools, followed a levy flight whose characteristics were correlated with the degree of clustering of their target fish. Levy mediated walks have been found in the tracks of diverse animal species, including soil amoebas (Levandowsky *et al.* , 1997), fruit flies (Cole, 1995), albatross (Viswanathan *et al.* , 1996), jackals (Atkinson *et al.* , 2002), and elephants (Xiaohua *et al.* , 2007).

Investigators wishing to assess the directedness of a walker have used the concept of fractal dimension (e.g. Fritz *et al.* , 2003). A *fractal* is "a rough or fragmented geometric shape that can be subdivided in parts, each of which is (at least approximately) a reduced-size copy of the whole" (Mandelbrot, 1983). This property is termed *scale independence*. The *fractal dimension* is a measure of the ability of the fractal to fill space. A straight line has a fractal dimension of 1 (taken to be the most directed form of walk), whereas a purely random (undirected) walk has a fractal dimension equal to its spatial dimension (such as 2 or 3). Turchin (1996) criticizes this method on the basis that animal tracks are usually better described by a correlated random walk, not a purely random walk, so that the underlying assumption of scale independence is violated and fractal methods cannot be meaningfully applied. Benhamou (2004) agreed and suggested a sinuosity index for application to search paths.

Kareiva & Shigesada (1983) present a method for testing whether individuals are moving according to a correlated random walk, using frequency distributions of observed step lengths and turning angles as well as observed net squared displacement. They derived a general model for a correlated random walk, which yields an expected net squared displacement as a function of the number of steps taken. Investigators use their observed tracking data to form frequency distributions of step lengths and turning angles from which the method predicts expected net squared

displacements for their study population. Sufficient deviation of the observed net squared displacement from this expectation indicates that the study organisms are moving according to a process other than a correlated random walk. Although this work has been widely applied to insects (e.g. Root & Kareiva, 1984) as originally intended, it has more general applicability, for example it has been applied to caribou (Bergman *et al.* , 2000), seals (Austin *et al.* , 2004), and elephant herds (Xiaohua *et al.* , 2007). Barrett & Lowen (1998), building on the work of Waser (1976), have done similar work for random walks that include barriers.

Track information has been used to define *home range*, the area or volume over which an animal or population normally travels in pursuit of its routine activities (Okubo & Levin, 2001). At its simplest, the home range is given by a polygon enclosing all available location estimates for an individual or a population (White & Garrott, 2006). A more useful concept than that of a bounded area of activity is the *utilisation distribution* introduced by Jennrich & Turner (1969), which gives the proportion of their time that an animal or group of animals spend within a region of space (Matthiopoulos, 2003). This idea results in a contoured map instead of a single polygon. This too can be estimated in a very direct, empirical way by placing a grid over the area of interest, interpolating the track between location fixes, and summing the number of tracks that cross each cell (McConnell *et al.* , 1999). Alternatively, kernel density methods may be used (Wood *et al.* , 2000; Worton, 1989; Matthiopoulos, 2003). Here a form of density distribution (the kernel), such as a Gaussian, is selected, and each location is considered to be at the centre of a distribution of this form. A spread parameter, such as the variance in the case of a Gaussian kernel, must also be specified. The sum of all the overlapping distributions is taken to represent the utility distribution. More sophisticated kernel methods use environmental data to tune the spread parameters of the individual distributions (Worton, 1989; Matthiopoulos, 2003; De Bie & Cristianini, 2004; Muller *et al.* , 2001). More so than home range, the concept of the utility distribution inherently incorporates aspects of animal behaviour. Therefore, where home range methods tend to be very empirical, utility distributions are more often estimated using models that take account of environmental factors (such as habitat preference) that may influence the animal's choice of location. Indeed, habitat preference alone may be the focus of the study (Mauritzen *et al.* , 2003).

Another area of research that may incorporate aspects of movement behaviour is that of location estimation. Track data collected through direct observation, radio-tracking, and satellite tracking give relatively precise estimates of location but archival tags do not give direct position estimates. Instead, their time-stamped

measures of ambient light are used to estimate the times of sunrise, sunset, and noon and consequently latitude and, with less precision, longitude, using a computer model. In addition to light, archival tags typically measure temperature and pressure (a proxy for depth). The temperature measure can be used along with satellite remotely sensed ocean temperatures or model-derived temperatures at depth to further refine these model estimated location estimates (Teo *et al.* , 2004; Royer *et al.* , 2005). Sumner *et al.* (n.d.) used knowledge of elephant seal foraging behaviour and swimming speed to refine geolocation estimates for elephant seals foraging in the Antarctic. Their model provides probability distributions of likely location. Sibert *et al.* (2003) and Sibert *et al.* (2006) used a state space model to simultaneously estimate location, tag shedding position, and movement parameters from archival tag data for tunas.

2.4.2 Random walks and advection-diffusion models

A random walk model represents a path as a series of discrete steps between successive locations. They may be parameterized either in terms of the successive locations, or in terms of the successive steps. In the first case, locations will be drawn from a distribution such as a normal or a Levy, and in the second, step lengths will be drawn from one distribution, such as a normal, and turning angles will be draw from another, such as a wrapped Cauchy or Weibull distribution (e.g. Morales *et al.* , 2004; Zollner & Lima, 1999). Many models of movement behaviour include some form of random walk component.

The continuous limit of a pure random walk is a simple diffusion model (Okubo, 1980), which is described using a differential equation. Non-random, directed movement can be achieved by adding an advection term, yielding an advection-diffusion model (Sibert, 1984). Other aspects of behaviour, such as a tendency for individuals to aggregate, can also be added (Mullen, 1989) but these require more sophisticated systems of differential equations (see Okubo, 1980; Okubo & Levin, 2001) which are correspondingly more difficult to solve (Dalgaard & Larsen, 1990). These models have been widely used (Skellam, 1951; Beverton & Holt, 1957; Jones, 1976; Grunbaum, 1999, 2000; Sibert *et al.* , 2003; Crittenden, 1994; Maury *et al.* , 2001; Maury & Gascuel, 1999; Sibert *et al.* , 1999) and can incorporate environmental data such as habitat type (Ovaskainen, 2004), but are difficult to work with and are not as easy to build in a modular way as are models that use difference rather than differential equations. The solution of the resulting system of partial differential equations quickly becomes algebraically intractable (Grunbaum, 2000), placing limitations on

the complexity of the behaviour that can be modelled in practice.

Advection-diffusion models are of most value when modelling systems in which movement is most strongly influenced by passive drift, such as larvae moving in ocean currents (e.g. Kasai *et al.* , 1992; Walters *et al.* , 1992), rather than by active swimming. In fact, Yamamura *et al.* (2003) argues against using diffusion models on the grounds that random walk models achieve the same result but have more biologically meaningful parameters than the difficult to interpret diffusion coefficient.

2.4.3 Bulk transfer process

Models that simulate the path taken by a moving individual often use a random walk and may therefore be called a 'random walk model', or may go by some other name when the random walk forms only a minor part of the whole. Similarly, models that do not simulate the path but simply allow organisms to move between pre-defined areas (boxes), according to a matrix of transition probabilities, are said to use a bulk transfer process (Sibert & Fournier, 1994) or are called box models (Porch, 1995). Confusingly, this is often termed a Markov process (Myers *et al.* , 1997) and the models that use them are called Markov process models (Matis *et al.* , 1992). Unfortunately, the term 'Markov process' could equally be applied to a random walk.

If the population is modelled as a group, proportions of the whole will be redistributed among areas at each time step according to the transition matrix. Alternatively, if individuals are modelled, a random number will be drawn for each individual, at each time step, and the transition matrix will be used to decide whether, and where, each individual will move. While random walks are suitable for modelling track data, bulk transfer processes lend themselves to tag-recapture data. They are most appropriate when the focus of the investigation is on the movements between metapopulations (e.g. Matthiopoulos, 2005).

The continuous (in time but not space) analogue for this type of model is one in which a movement rate parameter replaces the probability of moving within a time step (Polacheck, 1990).

2.4.4 IBMs

At the other end of the spectrum from differential equation models are the individual- or agent-based models, which are characterized by discrete variables, flexibility, and

ease of construction. Each individual in a population is modelled separately, and aspects of its life are modelled as events that occur with a particular probability. Random numbers are generated so that Bernoulli trials can be conducted for each process in the life of each individual. Movement may be modelled as a random walk, with probability distributions governing steps, or as a bulk transfer process, with a transition matrix giving the probability of moving from one area (or box) to any other (Myers *et al.* , 1997).

IBMs are tremendously flexible and easy to construct so that almost any degree of realism can be attained (DeAngelis & Gross, 1992; Grimm, 1999; Werner *et al.* , 2001; Bian, 2003). Several IBMs can be joined in an interacting hierarchy (Palmer, 1992). Their flexibility and almost infinite capacity for complexity is both their strength and their downfall because a model that too closely replicates reality does not provide useful generalizations about the system being modelled (Grimm, 1999). It is not clear how to estimate the parameters of IBMs. Typically, comparison of IBM results and observations is by eye (e.g. Railsback, 2001; Wiktorsson *et al.* , 2004; Ollason *et al.* , 1997), by comparing summary statistics (e.g. Haas *et al.* , 2004) or uses a statistical goodness-of-fit test such as a Chi-squared (French *et al.* , 1989). None of these methods easily lends itself to the iterative process of model fitting whereby parameter values are selected to maximize the fit between model and data. The parameter values and probability distributions used by IBMs are usually derived directly from data, however, inference is usually forgone and these parameter values are held fixed (e.g. Broekhuizen *et al.* , 2003; Bartsch *et al.* , 2004; Railsback *et al.* , 2003). Inference for IBMs has been achieved using neural networks and the genetic algorithm (Huse *et al.* , 1999; Huse, 2001; Wang & McKenzie, 1999; Eiane & Parisi, 2001). However, this process somewhat negates the advantage of IBMs that they are easy to use, and the parameters of neural networks seldom have biological meaning.

Given their realistic complexity, IBMs do offer promise as operating models, from which 'data' is simulated that may be used as 'observations' for simulation testing of other statistical models (Xiao, 2000). In addition to ecology, IBMs are used in the fields of computer simulation (for games), robotics (to control the movements of autonomous individuals) (Hostetler & Kearney, 2002; Wang & McKenzie, 1999), and have been used in modelling the stock market (Farmer, 2001).

2.4.5 Time series models and state space models

A time series is a collection of observations made sequentially through time (Chatfield, 2004). Time series models specifically account for the inherent autocorrelation in this sequence. Just as the advection-diffusion model can be seen as the continuous analogue of a discrete random walk, so the continuous analogue of a time series model is the stochastic differential equation model (Harvey, 1990). The overlapping nature of all the models discussed here is evident in the fact that the simple advection model is a special case of the stochastic differential equation model (Preisler *et al.* , 2004). Again, the continuous model is more difficult to solve and less flexible to work with than the discrete time models (McDonald & Sandal, 1999). Nevertheless, stochastic differential equation models have been applied to animal track data (Brillinger *et al.* , 2002), which invariably have unequal time steps.

Classical time series analysis concentrated on the modelling of stationary time series (those whose mean does not change over time), the autoregressive integrated moving average (ARIMA) models of Box & Jenkins (1976) being dominant. Non-stationary time series (whose mean exhibits a trend over time) can be made stationary by differencing - this is the precept of the ARIMA models (Chatfield, 2004), however, differencing does not always produce a stationary time series nor is stationarity always of interest. Non-stationary time series can be modelled using structural models, which decompose the time series into trend, seasonality and irregular variation (Chatfield, 2004). This idea is developed further in the state space model (SSM) (Harvey, 1990) - a flexible, general framework for modelling time series that can encompass most ARIMA models as special cases (Durbin & Koopman, 2001). Although traditional time series and state space models can be regarded as equivalent (Aoki, 1987), the emphasis of SSMs is on modelling trend rather than on reduction to stationarity.

An attractive feature of state space models is their ability to estimate both process and observation errors, and to separately estimate the variances of each. Traditional models incorporate only one source of error or, if two, require that one of these, or their ratio, be fixed. The state space framework is a flexible one, allowing realistic incorporation of behaviours and external variables that might influence behaviour. The very natural way in which SSMs allow for autocorrelation lends them to application to track data. For these reasons, the state space modelling framework shows promise for modelling movement behaviour. It has been applied to the behaviour of fishing vessels (Dorn, 2001) and to satellite track data for moving animals (both real and simulated) (Jonsen *et al.* , 2003, 2005, 2006; Flemming *et al.* , 2006; Sibert *et al.* , 2003) as well as to location estimation from archival tag data (Sibert *et al.*

, 2003; Sumner *et al.* , n.d.). Section 2.5 discusses the SSM framework in greater detail and shows how it has been adopted, and is still being developed, by natural resource modellers.

2.4.6 Discrete states and hidden Markov models

Models of behaviour sometimes recognize discrete behaviour states such as 'feeding', 'travelling', and 'searching' and allow the parameters of the model governing behaviour to differ for each state. In this way Morales *et al.* (2004) modelled movement behaviour as a set of random walks and Jonsen *et al.* (2005) as a switching state space model.

Both Morales *et al.* (2004) and Jonsen *et al.* (2005) were interested in modelling the path of the moving animal but, in other cases, the focus of the investigation has been on identification of these discrete behavioural states (e.g. Franke *et al.* , 2004, 2006). In such cases an appropriate modelling framework is a discrete time series model, in which the state variable takes on discrete values (MacDonald & Zucchini, 1997). Of these, the focus of this section will be on the hidden Markov model (HMM) (Rabiner, 1989), which has been used to model movement behaviour (Franke *et al.* , 2004, 2006).

Just as bulk transfer models use a transition matrix to govern the probability of moving from one area into another, HMMs use a transition matrix to govern the probability that the state variable will switch between discrete states.

Franke *et al.* (2004) present an HMM for caribou, which they applied to track data in the form of a sequence of step lengths and turning angles. They pre-specified the number of behavioural states (three) but made no assumptions about the parameter values for each state. The modelling process allocated each step to a behavioural state and estimated the transition matrix. This allowed the investigators to re-examine the data and to speculate on the behaviours associated with the three states (they called them 'bedding', 'foraging' and 'relocating'). By applying the model to data from different individuals and comparing the results, they concluded that individuals used the same areas in different ways. By comparing the associations between states and the habitats in which those occurred, the authors were able to speculate on different habitat usages.

Franke *et al.* (2006) used a similar HMM applied to wolf satellite tracks in the

form of step length, turning angle and travel rate to identify kill sites. They used either a 2-state or a 3-state model, both of which identified one state that represented feeding on a carcass. They were able to compare their model-predictions of kill sites with known kill-sites identified during twice-daily aerial surveys and found that the models correctly identified 74-77% of kill sites with those of the largest prey, caribou, most likely to be correctly identified because those carcasses took longest to consume.

In companion papers Guilford *et al.* (2004) and Roberts *et al.* (2004) present an HMM for pigeon navigation applied to bird flight tracks in the form of a sequence of track complexity measures. They conclude that three distinct behavioural states exist. These correspond to high-complexity behaviour immediately following release when the birds are most disoriented; medium-complexity (medium entropy) behaviour during which birds do not alter their course unless reasonably sure of the direction to their loft; and low-complexity (high entropy) behaviour when birds are close enough to their loft to be sure of its position.

Discrete state models provide insights into movement behaviour that could inform subsequent model-building of models that explicitly simulate movement paths. Alternatively, state switching can be incorporated into state space models of movement behaviour (Morales *et al.* , 2004; Jonsen *et al.* , 2005; Fruhwirth-Schnatter, 2001; Kim & Nelson, 1999).

2.5 SSMs in more detail

2.5.1 Nomenclature

It is perhaps not surprising that a modelling framework as flexible and therefore as broadly encompassing as the state space model should be presented in the literature with somewhat confused terminology. Harvey (1990), who did not originate these models but who did much to advance their use, uses the term 'state space model'. Many models can be expressed in state space form. Buckland *et al.* (2004) points out that even IBMs can be expressed as SSMs (more of this below), so that names such as linear trend model, and structural time series model (Freeman & Kirkwood, 1995) or (nonlinear) structural equation model (Lee & Song, 2004; Lee & Tang, 2006) can be as validly applied to certain formulations as can the term state space model (Schnute, 1994). Borchers *et al.* (2002) used the term 'state model',

Newman & Lindley (2006) used 'hidden process model' and Buckland *et al.* (2004) and Thomas *et al.* (2005) used 'state space model' even though all work in the same organisation. Adding to this confusion is the work of West & Harrison (1997), who present a Bayesian approach (Harvey (1990) mainly worked with maximum likelihood) to that subset of SSMs whose coefficients update with time. (Such a model is presented in Chapter 4.) They term these linear, or nonlinear, dynamic models. Further, some authors appear to reserve the term SSM for linear, Gaussian forms, although this is not done by Harvey (1990). The term 'state space model' is used in this thesis and is applied in its most general sense to encompass nonlinear, non-Gaussian, forms and those whose parameters update with time.

2.5.2 Model framework

The following description of the state space modelling (SSM) framework uses the notation of Schnute (1994) who presents a generic state space framework for models that are not necessarily either linear or Gaussian.

State space models represent the progress through time of an unobserved state vector X whose value X_t during time step t is a first-order Markov function $F(\cdot)$ of its value in the previous time step X_{t-1}, a vector of control variables Z_t, and the model parameters θ, with stochastic error δ_t (the process error)

$$X_t = F(X_{t-1}, Z_t, \theta) + \delta_t. \tag{2.1}$$

The state equation could take the form of a structural time series, consisting of any number of layers. Even an IBM could be written in SSM form by adding a separate line to the state equation for each individual (Buckland *et al.* , 2004).

Observations Y_t made during time step t are related to the state variable X_t though some function $G(\cdot)$ of the parameters θ, and possibly the control vector Z_t, with stochastic error ϵ_t

$$Y_t = G(X_t, Z_t, \theta) + \epsilon_t. \tag{2.2}$$

The control vector Z represents observed data that, unlike the observations Y, do

not depend directly on the states X_t. The parameters of the distributions of the random errors δ_t and ϵ_t (which are not necessarily normal) are model parameters and are thus part of the vector θ. These errors are assumed to be independent of one another, and between time steps.

These two equations, together with the parameters and distributional forms of the error terms, and an assumption regarding the initial condition of the state variable, complete the specification of the state space model (Harvey, 1990). If the model equations are linear and the errors are assumed to be normally distributed then the model is termed a linear, normal SSM.

Likelihoods

The *likelihood* function $L(Y \mid Z, \theta)$ provides a deductive way of comparing the probability of the data Y under different values of the parameters θ (and given the control variable Z) (Pawitan, 2001). This definition is common to both Bayesian and classical statistics. Bayesians and classicists differ, however, in their treatment of parameters and latent data and therefore in their view of alternative quantities, perhaps incorrectly termed 'likelihoods', that may be used for inference. These are described below. In this context the likelihood itself has been termed the *true likelihood*, for clarity.

The control variable Z, often omitted from discussions regarding likelihoods (e.g. de Valpine, 2002), is retained here because it is pertinent to the modelling framework presented in Chapters 3 - 5.

State space models consider latent data or unobserved states X in addition to parameters θ. To a classicist, the latent data are random variables or nuisance parameters that must be integrated out before inference can be made regarding the model parameters θ, which are assumed to have some fixed, true value. To a Bayesian, both parameters and latent data are random variables and it is valid to make inferences about both. In the Bayesian context an SSM is a hierarchical model because the unobserved states (hyperparameters), being random variables, are functions of other random variables (the parameters).

The joint distribution $p(Y, X \mid Z, \theta)$ of the observed data Y and the latent data or unobserved states X, given the parameters θ and the control data Z, can be expressed in augmented form

$$p(Y, X \mid Z, \theta) = p(Y \mid X, Z, \theta)\ p(X \mid Z, \theta). \tag{2.3}$$

This quantity has been termed the *hierarchical likelihood* (Pawitan, 2001, Section 16.3), the *errors in variables (EV) likelihood* (de Valpine, 2002), the *penalized likelihood* (Millar & Meyer, 2000b), and the *complete data likelihood* (Robert & Casella, 1999). The term 'complete data likelihood' is the term preferred in this thesis because it recognizes that equation 2.3 is the true likelihood if both X and Y are observed.

Because X is not observed, the *true likelihood* (for both Bayesians and classicists) is given by equation 2.3 integrated across the latent data or unknown states X, which is the same as integrating across the process errors (de Valpine, 2002)

$$L(Y \mid Z, \theta) = \int p(Y \mid X, Z, \theta)\ p(X \mid Z, \theta)\, dX. \tag{2.4}$$

For state space models this is usually a high dimensional integral, and is thus the cause of difficulty regarding inference (Buckland *et al.* , 2004).

In Bayesian terms, the probability $p(X \mid Z, \theta)$ (equation 2.3) of the latent data or unobserved states X given the control data Z and parameters θ can be regarded as a prior for X, with respect to a likelihood $p(Y \mid X, Z, \theta)$ for the data Y given the states X, the control vector Z, and the parameters θ (Schnute, 1994). To distinguish this from other densities, in this thesis $p(Y \mid X, Z, \theta)$ is termed the *full likelihood.*

For a Bayesian, it is valid to use the complete data likelihood for inference, particularly if the values of the states X are of interest (Schnute, 1994). However, de Valpine (2002) indicates that while the true likelihood has desirable asymptotic statistical properties, the complete data likelihood may lack these. This likelihood cannot be used for classical maximum likelihood estimation without first specifying the value of one of, or the ratio between, the error variance components of θ (for the errors in the observations Y and the latent data X). This is unnecessary in a Bayesian estimation of the joint posterior for all model parameters (and perhaps latent data X).

While de Valpine (2002) indicates that the complete data likelihood may lack the

desirable asymptotic qualities of the true likelihood, he concedes that in typical fisheries modelling cases asymptotic properties may have little relevance due to paucity of data and that the greater uncertainty is in how to adequately represent a complicated system as a set of equations. The same argument would apply, if anything more strongly, to modelling of movement behaviour where it is far from clear what an adequate representation would be, or even what the components are, of behaviour.

The Monte Carlo approach to the problem of calculating the high dimensional integral for the true likelihood recognizes that equation 2.4 indicates that the true likelihood $L(Y \mid Z, \theta)$ is the expectation $E[p(Y \mid X, Z, \theta)]$ with respect to $p(X \mid Z, \theta)$. Thus if $X^{(i)}$, $i = 1 \ldots N$ represents an independent set of draws of X from $p(X \mid Z, \theta)$ then

$$\hat{L}(Y \mid Z, \theta) = \frac{1}{N} \sum_{i=1}^{N} p(Y \mid X^{(i)}, Z, \theta) \tag{2.5}$$

is the Monte Carlo estimate of the true likelihood. This is termed the *simulated likelihood* by Millar (2004) (although it is not a likelihood but rather an estimator of the true likelihood) who uses a multivariate normal approximation to $p(X \mid Z, \theta)$ to estimate the true likelihood by importance sampling. Although the variance associated with this estimator can be unacceptably large, Millar (2004) was able to successfully apply it to several models including an SSM.

2.5.3 Hierarchical SSMs

The term hierarchical SSM (noting that a Bayesian SSM is already a hierarchical model) is used to describe and SSM whose parameters are drawn from hyperdistributions defined by hyperparameters (Jonsen *et al.* , 2006)[][e.g.]. Here, a separate model is fitted to each individual from which observations were made, so that each has its own parameters. The parameters for all individuals are considered to be jointly distributed - the hyperprior. The hyperprior and its (hyper)parameters form a higher model layer, thus the term 'hierarchical'.

The constraint imposed by the hyperprior restricts the number of degrees of freedom introduced by the additional parameters. In this way separate (but jointly distributed) parameters may be estimated for several individuals whose behaviour

is governed by the same model (Jonsen *et al.* , 2006; Dorn, 2001; Newman, 2000; Rivot *et al.* , 2004; Newman & Lindley, 2006). Environmental covariates can be included in the hierarchical layer, as is recommended by Newman & Lindley (2006). A general framework for hierarchical SSMs is presented by Gamerman & Migon (1993) and suitable hyperpriors for variance parameters are discussed by Gelman (2006).

2.5.4 Inference for SSMs

The main stumbling block to the widespread implementation of SSMs has been difficulty in inference (Buckland *et al.* , 2004). Either, the likelihood must be integrated across all possible values of the state variables, and then the parameter values estimated, or the high dimensional problem of estimating both parameter and state values must be solved. The earliest hope for a solution to this problem was the Kalman filter (Schweppe, 1965; Kalman, 1960; Kalman & Bucy, 1961; Meinhold & Singpurwalla, 1983), a recursive algorithm which provides maximum likelihood estimates of the state of the system as well as of the error variances (Harvey, 1990). Unfortunately, the Kalman filter is only applicable to linear, Gaussian SSMs. The advance of SSMs has also been dependent on improving computer technology, with the use of even the relatively straightforward Kalman filter possible only once sufficient computing power had become available for its implementation for high dimensional models (Harvey, 1990).

Kalman filter

The Kalman filter allowed modellers to recast existing models, which typically included only one source of error, in the state space form using both process and observation errors. The greater realism afforded by this approach was attractive and it was shown that the results of fisheries population dynamics models could differ significantly when recast in this fashion (Schnute, 1991; Kimura *et al.* , 1996). For this reason linear, normal SSMs using the Kalman filter were adopted and applied to the simpler stock assessment models such as the production or biomass dynamics (Freeman & Kirkwood, 1995; Reed & Simons, 1996) and delay-difference models (Kimura *et al.* , 1996) and even to simple age-structured models (Sullivan, 1992). Movement was also incorporated into some SSM population models (Newman, 1998, 2000) although these simulate bulk transfer, not movement paths.

Realistic applications in biology typically require nonlinear equations (Anderson-Sprecher & Ledolter, 1991; Gautestad & Mysterud, 2005; Millar & Meyer, 2000b; Thomas *et al.* , 2005; Trenkel *et al.* , 2000) and t-distributed error forms provide more robust inference (Jonsen *et al.* , 2005). Attempts have been made to apply adapted forms of the Kalman filter to nonlinear cases; the most cited examples being the extended Kalman filter (Anderson & Moore, 1979) and the unscented Kalman filter (Julier *et al.* , 1995; Julier & Uhlmann, 2004). Vachhani *et al.* (2006) presents a combination of the two, able to cope with constraints on parameter values. Although these approaches may work well in simple cases (Groennevik & Evensen, 2001) and have been applied in more complex cases (Sibert *et al.* , 2006) simulation has shown that they can be unacceptably inaccurate, particularly when the problem is highly nonlinear (Wang, 2007; Pella, 1993; Gudmundsson, 1994, 1995; Punt, 2003).

Monte Carlo methods

Widespread use of nonlinear, non-Gaussian SSMs had to await the next breakthrough - even more powerful computing technology and the effective but highly computer intensive Monte Carlo methods (Stewart, 1983; Carlin *et al.* , 1992; Buckland *et al.* , 2004; Newman & Lindley, 2006; Thomas *et al.* , 2005; Hobbs & Hilborn, 2006) that provide a means of numerically evaluating otherwise intractable distributions. A particular boost to this research has been the availability of the WinBUGS statistical program (Spiegelhalter *et al.* , 2004; Meyer & Millar, 1999a), which has made SSMs more accessible, particularly to biological modellers (Rivot *et al.* , 2004; Jonsen *et al.* , 2006; Morales *et al.* , 2004; Thompson *et al.* , 2005).

Monte Carlo methods circumvent the need to analytically evaluate the likelihood, and posterior, and consequently free the investigator to consider nonlinear, non-Gaussian forms. These are discussed in greater detail in Section 2.6. .

Other methods

Other methods of inference, beside those discussed above, may hold promise for use with SSMs. de Valpine & Hastings (2002) and de Valpine (2002) present NISS (numerically integrated state space method) for fitting state space models to time series of population abundances. Based on the algorithm of Kitigawa (1987) the method uses the logic of the Kalman filter but because it is implemented numeri-

cally, can accommodate nonlinear and non-Gaussian models. This is a non-Bayesian method, using maximum likelihood estimation. NISS has been shown to perform better than the Kalman filter when applied to a linearized population dynamics model that included both observation and process errors. Nevertheless it gave quite biased estimates for some simulations and Punt (2003) recommended that future studies attempt Monte Carlo methods instead.

Kalman filtering and MCMC are not necessarily mutually exclusive, Carter & Kohn (1994) used Gibbs sampling to carry out Bayesian inference on a linear state space model but used a Kalman filter to generate the unobserved states.

2.5.5 SSMs in wildlife population and movement modelling

Nonlinear, non-Gaussian SSMs are now available to natural resource modellers. Schnute (1994) presents a statistical SSM framework that encompasses most models used in fisheries science and Buckland *et al.* (2004) and Thomas *et al.* (2005) show how the SSM framework can encompass most wildlife population dynamics models previously written as matrix population models (see Caswell, 2001). An application of this framework to salmon is given by Newman & Lindley (2006). Meyer & Millar (1999a) developed a nonlinear SSM form of the traditional fisheries delay-difference model, and more complex age-structured models were produced (Schnute, 1991; Millar & Meyer, 2000a).

Where diffusion models treat movement as a continuous process in both time and space, SSMs treat time as discrete, but are flexible with regard to space. Early SSMs applied to animal movement problems used a small number of large areas and a bulk transfer approach for moving animals between these areas (Newman, 1998, 2000). Tag-recapture data were used and although more detailed movement models are now possible, this method is still applicable particularly when applied to tag-recaptures (Harrison *et al.* , 2006). Increases in computing power and the availability of satellite tracking data have resulted in the development of continuous space SSMs (Jonsen *et al.* , 2003, 2005, 2006; Flemming *et al.* , 2006).

An advantage of the separation of observation errors from process errors in the context of modelling satellite tracks is that this facilitates modelling of the unusual error structure of GPS location data, separately from the process errors likely to

result from incomplete understanding of highly complex behaviours. Observation errors in longitude are typically greater than in latitude (Sibert *et al.* , 2003; Vincent *et al.* , 2002), whereas there is no reason to assume different variances for latitude and longitude in the process errors. As mentioned in Chapter 1, when Argos provides a sequence of locations, it also allocates each location to one of six quality grades. Vincent *et al.* (2002) collected a dataset of known and corresponding Argos estimated locations from which Jonsen *et al.* (2005) estimated t-distributed errors separately for latitude and for longitude. There is no reason to assume such categorisation for process errors, which one would be forced to do if modelling both observation and process errors as a single error source.

2.6 Monte Carlo methods

For distributions that are analytically intractable, or whose analytical form cannot be obtained, Monte Carlo methods provide a means of making draws from that distribution. From these samples, properties of the distribution and expectations with respect to that distribution can be estimated (Gilks *et al.* , 1996). These methods are commonly used in Bayesian analyses because of the need to estimate the posterior distribution, or at least quantities relating to it, such as its mode and marginal distributions for the parameters. Nevertheless, Monte Carlo methods are not themselves either Bayesian or classicist.

Commonly used Monte Carlo methods include the importance samplers (Gelfand & Smith, 1990) and the Markov chain Monte Carlo (MCMC) methods, in particular Gibbs sampling (Geman & Geman, 1984; Tanner & Wong, 1987; Gelfand & Smith, 1990), and the Metropolis-Hastings algorithm (Metropolis *et al.* , 1953; Hastings, 1970). All of these methods have many variants, for more detailed descriptions see Gilks *et al.* (1996) and Chib & Greenberg (1996).

2.6.1 Importance samplers

Importance samplers bear some resemblance to the Metropolis-Hastings algorithm in that parameters are sampled from a distribution that is not the distribution of interest but that is chosen for its similarity to that distribution and for its ease of sampling. Parameter values are sampled from the importance distribution and are then weighted by the ratio of the densities of the sampling distribution and the

distribution of interest, evaluated at sample. Importance sampling methods can be made more efficient by using the accumulated sample to improve the importance distribution during the execution of the algorithm. Examples include sampling importance re-sampling (SIR) (Rubin, 1987, 1988; McAllister & Ianelli, 1997), adaptive importance sampling (Evans, 1988; Kloek & van Dijk, 1978; West, 1993) and local adaptive importance sampling (Givens & Raftery, 1996), and annealed importance sampling (Neal, 2001).

A particular subset of importance sampler that has shown promise for application to time series models, the sequential importance samplers (SIS) (Liu & Chen, 1998; Liu & West, 2001) have been used by Trenkel *et al.* (2000), Buckland *et al.* (2004), Thomas *et al.* (2005), and Newman & Lindley (2006). These avoid the high dimensional integral of the SSM likelihood by splitting it into a sequence of lower dimensional integrals, one for each time step. The inherent time sequence of the SSM lends itself to this method. First, a large number of samples are drawn from the priors for the parameters and the initial states. Each set of parameter and initial state values, called a 'particle', is stochastically projected one time step forward using the state equation. The observation equation is then applied to these state values and the corresponding density distribution of these 'observations' is used to weight the original particles. Particles are then re-sampled from this weighted density. The process is repeated for each time step and the set of particles remaining at the end of the time-period provides an estimate of the posterior density given all the data (Thomas *et al.* , 2005). The chief difficulty encountered when using this method is particle depletion, whereby particles with small weights are unlikely to survive to the end of the time-period and those with large weights can be oversampled. Buckland *et al.* (2004) and Thomas *et al.* (2005) attempt to overcome this problem by using SIS with re-sampling and kernel smoothing (SISR/KS). However, this led to a new problem, that of having to select an appropriate smoothing parameter for the kernel.

2.6.2 Markov Chain Monte Carlo

The versatile MCMC methods are used in this thesis - detailed descriptions of these applications are given in Chapters 3 and 5. This section gives a broad overview of MCMC methods, and Section 2.6.3 discusses methods of investigating convergence of MCMC sampling chains.

MCMC methods provide a set of draws from a distribution of interest (usually the likelihood or posterior), which can be used to calculate quantities of interest

relating to that distribution, a simple example being the mode of the distribution. Each draw is likely to consist of a number of quantities, such as parameter values, which may be grouped into blocks. For the remainder of this section it will be assumed that the distribution of interest is a joint distribution for the model parameters (where the term 'parameter' includes all estimated quantities including unobserved data or states).

A popular MCMC method, the Gibbs sampler, requires that the full conditional distributions be known for all parameter blocks. The investigator selects a set of parameter values from the parameter space to serve as a starting point. A value is then drawn randomly from the conditional distribution for one of the parameter blocks, using the starting values for all the remaining parameters. This draw is stored, and is used to make a draw from the conditional distribution for one of the other parameter blocks. That value, in turn, replaces the starting value for the second parameter block and sampling proceeds until a draw has been made for every parameter block, yielding a new parameter set. This constitutes a single iteration of the Gibbs sampler. The new set of values is used to perform another iteration, yielding another set of parameter values - thus sampling proceeds as a Markov chain. Sampling proceeds until a large number of parameter sets have been drawn from the distribution of interest, and the investigator is satisfied that convergence has been achieved. Measures of convergence are discussed in Section 2.6.3.

Another popular MCMC method, the Metropolis-Hastings sampler, may be used when conditional distributions are not available for one or more parameter blocks - a proposal distribution is used instead. Greater efficiency is achieved by closer resemblance of this distribution to the distribution of interest. Again, a starting set of parameter values is chosen by the investigator and a draw is made for one of the parameter blocks, this time from the proposal distribution. Unlike Gibbs sampling, this candidate draw is not automatically accepted. A Bernoulli trial is performed, and the candidate value is accepted with a probability equal to the ratio of the distribution of interest to the density of the proposal distribution, evaluated at the candidate location. If the ratio exceeds 1, the candidate is always accepted. Therefore, draws taken from regions of parameter space where the proposal distribution has greater density relative to the distribution of interest, will be accepted less often, and those from regions where the proposal distribution has low density will be accepted more often. Draws from high density regions of the proposal distribution will be made too often, relative to the distribution of interest, but the lower acceptance rate compensates for this. Draws from low density regions of the proposal distribution, made too seldom, are correspondingly accepted more often.

A chain generated in this fashion converges asymptotically on the distribution of interest (Gilks *et al.* , 1996; Gamerman & Lopes, 2006).

Gibbs sampling is a special case of the Metropolis-Hastings sampler, and these have been combined as a Metropolis-Hastings within Gibbs sampler (Geweke & Tanizaki, 2001) and applied to SSMs (Gamerman, 1998; Millar & Meyer, 2000b; Lee & Song, 2004; Lee & Tang, 2006; Wang, 2007). This is the methodology used by the WinBUGS software (Spiegelhalter *et al.* , 2004) and its predecessor BUGS, which primarily implement Gibbs sampling but can also insert a Metropolis-Hastings step (Spiegelhalter *et al.* , 2004).

Typically, those who have used MCMC methods to solve high dimensional SSM problems have not attempted SIS and those who have concentrated on SIS have not yet explored MCMC methods (Buckland *et al.* , 2004; Thomas *et al.* , 2005). A number of variants on all of these methods exist (see Chib & Greenberg (1996) for a comprehensive review of MCMC samplers) so that there may be potential for greater efficiency in SSM inference. To give some examples, MCMC methods may be used in conjunction with rejection sampling (Tierney, 1994) and also with stochastic versions of the EM algorithm (Celeux & Diebolt, 1985). Tanner & Wong (1987) introduced data augmentation to improve the efficiency of the Gibbs sampler. Haario *et al.* (2006) presents a method for improving the efficiency of the Metropolis-Hastings algorithm, combining two earlier ideas, delayed rejection and adaptive Metropolis, to produce a hybrid that can be more efficient than either alone.

MCMC has opened a road to nonlinear, non-Gaussian SSMs. Nevertheless, many challenges remain and inference for SSMs is still an area of active research (Buckland *et al.* , 2004; Newman & Lindley, 2006), even for normal linear Kalman filtered approaches (Dennis *et al.* , 2006).

2.6.3 Convergence

It can be seen from the description, above, of MCMC samplers that the influence of the starting set of parameter values, whose selection is left up to the investigator, may influence the early draws of parameter values, but that this influence can be expected to wane. For this reason a *burn-in* period is often recommended where early draws are discarded (Gilks *et al.* , 1996).

The Markov chain nature of these samplers results in correlation between subse-

quent draws so that it might be advisable to *thin* the chain by discarding all but the jth iteration (where j is some positive integer).

The set of draws made by the Monte Carlo methods described above will, asymptotically, resemble the distribution of interest: so-called *convergence*. By plotting the values of each parameter against iteration, a visual representation is available from which the influence of the starting value, the serial correlation of the chain, and to some degree the convergence, may be assessed by examining any trend or cyclicity in this *trace* plot. Ideally, no evidence of pattern or trend should be visible. It is also desirable to begin the chain again at a different starting point and to assess whether a similar trace is seen, once the influence of the starting point has reduced (Gelfand *et al.* , 1990; Sorensen & Gianola, 2002). A number of more formal estimates of convergence are available (see Gamerman & Lopes, 2006; Gilks *et al.* , 1996; Robert & Casella, 1999; Cowles & Carlin, 1996; Raftery & Lewis, 1992; Gelman & Rubin, 1992; Geweke, 1992).

Several of these diagnostics have, conveniently, been implemented in the CODA software available through WinBUGS and R (Best *et al.* , 1996). The Geweke diagnostic assesses convergence of the mean of the chain using time series techniques (Geweke, 1992). The Gelman and Ruben statistic uses two or more starting points (and therefore chains) to compare the within chain and between chain variances for each parameter (Gelman & Rubin, 1992). The Raftery and Lewis diagnostic, based on two-state Markov chain theory, uses a single chain to estimate the number of iterations to discard during 'burn-in' as well as a minimum number of iteration required for convergence (Raftery & Lewis, 1992). The Heidelberger and Welch diagnostic tests for convergence, and assesses 'burn-in' by testing the chain for stationarity, sequentially discarding the first 10% of the chain until stationarity is achieved or until 50% of the chain remains (Heidelberger & Welch, 1983). CODA also reports autocorrelation at a range of lags in the chain (for each variable) as well as reporting cross-correlations between variables or each chain. High autocorrelation and high cross-correlation indicate slow convergence (mixing). High cross-correlations may indicate a need for reparameterization (Best *et al.* , 1996).

Ideally, convergence is assessed for every element in the parameter space, that is, for every scalar parameter and for every element of every parameter vector and matrix. For models that have a large number of parameters (such as those investigated in this thesis, which typically have in excess of 170 parameter and state elements) this task is not feasible and convergence is instead assessed by applying diagnostic tests to all scalar parameters and a random selection of elements of larger matrices, as

well as to quantities of interest such as the log-likelihood. Diagnostic tests were concentrated on a subset of the sensitivity tests to which the model was subjected and, in particular, to those simulation tests in which the estimation model most differed from the model used to simulate the data. The selection of the number of iterations to use to ensure convergence was deliberately generous.

2.7 Model comparison

2.7.1 Methods

The ability to assess whether the data are more consistent with one model than with another is an important aspect of inference. For models that are nested (the parameter set for one of the models is a subset of that used by the other) and that use maximum likelihood, the likelihood ratio test can be used to choose the model that best, and most parsimoniously, represents the data. In the context of realistic SSMs applied to movement data, this test will be available for only a small subset.

When models are not nested, or maximum likelihood is not used, alternative statistics to the likelihood ratio test must be used. The classical Bayesian means of comparison is the Bayes factor (Kass, 1993; Kass & Raftery, 1995; Lavine & Schervish, 1999). When only two models are compared, and the prior is a point hypothesis, the Bayes factor reduces to the likelihood ratio (Carlin & Louis, 2000). The Bayes factor is discussed in detail in Chapter 5.

The Bayes factor can be difficult to calculate, so that computationally easier alternatives are often used. The most common are Akaike's information criterion AIC (Akaike, 1973), Schwartz's Bayesian information criterion BIC (Schwarz, 1978), and a generalisation of the two, the deviance information criterion DIC (Spiegelhalter *et al.* , 2002; Berg *et al.* , 2004). The primary justification for using AIC is that it is asymptotically (with increasing sample size) equivalent to the Bayes factor. AIC is a function of the maximum likelihood (MLE) and the number of model parameters r

$$\mathrm{AIC} = -2\log(\mathrm{MLE}) + 2r. \tag{2.6}$$

AIC has been found to overestimate the optimal number of parameters needed by

the model (Carlin & Louis, 2000; Kadane & Lazar, 2004). BIC is preferred, being more parsimonious (Kass & Raftery, 1995; Kadane & Lazar, 2004). It differs from AIC only in that the 2 in the penalty term is replaced by a function of the sample size n

$$\text{BIC} = -2\log(\text{MLE}) + \log(n)\, r. \tag{2.7}$$

The model having the smallest AIC or the smallest BIC is the one given most evidence by the data. A popular variant on AIC, AIC_C corrects for overdispersion, see Burnham & Anderson (1998).

The sample size, or the number of parameters, may not be easily identifiable. For models that use a mixture of fixed and random effects, the number of parameters is not clear. Neither is it always clear what the effective sample size is, for example, when multiple measures were taken from each of a number of individuals. The DIC presents a solution to these problems. It does not require specification of the sample size, and it calculates a measure of the effective number of parameters. DIC, in its simplest form, is given by a function of the expected value of the likelihood $L(\theta)$ with respect to posterior $p(\theta \mid y)$, and the likelihood evaluated at the posterior mean $\bar{\theta}$

$$\text{DIC} = 2\, E_{\theta|y}[\, -2\log L(\theta)\,] + 2\log\, L(\bar{\theta}). \tag{2.8}$$

Again, it is the model corresponding with the lowest value that is considered to be most consistent with the data. Further information on the calculation of DIC is given in Chapter 5.

Different model selection methods applied in the same situation, might select different models so that the concept of the 'best' model must be defined in the context of the model selection statistic being used. Only in simulation, where data is generated using a model whose parameter values are known, and that model is one of the candidates from which the 'best' model is selected, can the concept of a 'true' model be used. Outside of simulation, it is highly unlikely that any of the candidate models will be an exact replica of the process that generated the data, so that the term 'best' model must be defined more loosely as a closer approximation to the true process. Further discussion and reviews on model selection methods can be found

in Kadane & Lazar (2004), Hoeting *et al.* (1999), Han & Carlin (2001), DiCiccio *et al.* (1997), and Gelfand & Dey (1994).

Rather than selecting the 'best' model, an approach that is more consistent with Bayesian thinking is to specify a prior distribution for all candidate models and then to estimate their joint posterior probability. This approach, Bayesian Model Averaging (BMA), is particularly appropriate when a model output is used to guide decision-making. BMA yields a probability measure for each possible value of that model output, which can be used in calculating the risk associated with a corresponding management action. Different models will each produce different posteriors for the value of the quantity of interest, a posterior for the models themselves can be used to weight those model posteriors, and consequently to produce a weighted average. BMA can be achieved in a number of ways, including by using Bayes factors as weights (Hoeting *et al.* , 1999). Alternatively, the parameter space can be expanded to include all model parameter spaces, and a model index can be introduced so that a joint posterior can be estimated. Such methods include the product space search method for Gibbs sampling (Carlin & Chib, 1995) and its Metropolized variant (Dellaportas *et al.* , 2002), reversible jump MCMC (RJMCMC) (Green, 1995), which uses a Metropolis-Hastings methodology, independence sampling Congdon (2006) and the hybrid Gibbs-RJMCMC sampler (Walker *et al.* , 2006). These methods can also be used to estimate the Bayes factor (Huelsenbeck *et al.* , 2004; Congdon, 2006; Han & Carlin, 2001).

In the field of movement modelling, inference is at an early stage. For many animal species, satellite tracking presents the first viable tracking method available. Therefore, little prior information is likely to be available for movement behaviour. Little is likely to be known regarding which external and internal cues might be influencing moving individuals, or how these individuals respond. The work of investigators is to sift hypotheses of movement behaviour, establishing which are most consistent with available tracking data. Model selection tools would therefore be of more value than model averaging, which is primarily a means of accurately quantifying uncertainty and therefore the risk associated with making a decision based on current knowledge. Conceivably, situations might exist where movement behaviour could be used to guide management decisions and model averaging might then be useful, for example, in selecting the location of a proposed marine protected area. However, the present study focuses on hypothesis testing and therefore on model selection.

2.7.2 Movement studies

The few published SSM studies of movement behaviour have tended to focus on model fitting rather than on model comparison. Jonsen *et al.* (2006), when presenting an SSM applied to satellite tracks for leatherback turtles, state that their future work will include some measure of model comparison but state that DIC cannot be applied to models such as theirs that use robust t-distributed observation errors. The t-distribution is not an exponential family, therefore the likelihood is not guaranteed to be log-convex. Thomas *et al.* (2005) present an SSM framework for application to wildlife population dynamics modelling and are mindful of future management decisions that may be based on their framework. They too have not yet included model comparison in their work but hope to do so in the future. They suggest using AIC to give a form of penalized likelihood weighting to alternative models for subsequent model averaging. Alternatively, they hope to extend their SIS algorithm so that it can perform Bayesian model averaging.

Newman (1998) used BIC with an SSM that modelled the movement of salmon between areas as a bulk transfer process. Wiktorsson *et al.* (2004) also used BIC, in their case with random walk models of insects moving within an enclosed arena. Harrison *et al.* (2006) used AIC with a Bayesian SSM that applied a bulk transfer process to seals moving between metapopulations; tag-recapture data were used.

The WinBUGs package (Spiegelhalter *et al.* , 2004) calculates the DIC statistic so that investigators who implement their models in WinBUGs typically use DIC for model comparison (Morales *et al.* , 2004). As WinBUGs appears to be a popular tool for implementation of SSMs in biology it is likely that DIC will be widely used. However, the caution raised by Jonsen *et al.* (2006) regarding t-distributed errors introducing log-concavity, should be noted.

Lee & Song (2004) and Lee & Tang (2006) used the Bayes factor in conjunction with an SSM. Rivot *et al.* (2004), too, used the Bayes factor however their calculation used the harmonic mean estimator, which Chapter 6 demonstrates should not be used, despite published assurance to the contrary (Kass & Raftery, 1995).

2.8 This study

The state space model has many aspects that make it suitable for modelling movement behaviour. It is a time series method, able to model the inherent autocorrelation in track data, it is flexible, allowing the relatively easy incorporation of a wide range of behaviours, it explicitly models both process and observation errors, and early problems of inference have largely been overcome. DIC is a model comparison method suitable for use with SSMs. Subsequent chapters show that it has been possible to develop a flexible state space modelling framework for representing a wide range of movement behaviours and that this has been used together with simulated tracks to successfully distinguish the behaviour that was used to simulate the track. This model was applied to track data from white sharks to investigate whether they navigate by following a line of bathymetry (isobath). The method showed that this was not the case. During this investigation a number of lessons were learned and avenues for future work uncovered, these are discussed in the final chapter.

Chapter 3

Structural time series modelling framework for animal movement behaviour

3.1 Abstract

A flexible Bayesian state space movement modelling framework has been developed for application to track data. The framework models observation and process errors separately, accounts for the autocorrelation inherent in track data, and accommodates irregular time steps between location observations. Movement behaviour is incorporated using one or more advection fields, which are calculated externally to the main model, using co-models. These co-models incorporate theories regarding how animals might be moving, and may use covariate data. An advection field indicates the direction in which an animal is inclined to move, depending on its present location. Advection fields can be combined, as a weighted sum, so that more than one motivation can be combined to provide an overall movement behaviour. The use of advection fields and the separation of the co-models from the main model provides great flexibility in the range of behaviours that can be modelled.

Track data often includes large gaps (in both time and distance) where no observations were made. Straight-line interpolation between such points may not be realistic, particularly if impassable barriers separate observed locations. To cater for this, the modelling framework allows the estimation of latent locations. Inertia, the directional persistence often evident in animal movements, is also modelled.

A model, selected from the family of models represented by this modelling framework, is applied to tracks simulated using advection fields representing either movement in a constant direction, or movement that changes direction depending on location. Constant advection fields allow the use of the Gibbs sampling MCMC routine whereas changing fields require that a Metropolis-Hastings step be used to estimate the true path taken by the animal. Simulations were conducted using either a single advection field, or two fields operating together.

Simulations showed that it is difficult to distinguish between inertia and advection, however, when these have different directions for at least part of the track, the model is capable of disentangling them. When more than one advection field is used together, the more they differ the less correlated their coefficients will be. It is also difficult to estimate the precision parameters for the observation and process errors, the model is guided by the prior for these quantities. However, incorrect estimation of these precision parameters does not appear to bias the remaining model parameters.

The effect of incorrect parameter estimation on model comparison is explored in Chapter 5.

3.2 Introduction

This chapter presents a flexible, state space movement modelling framework capable of representing the movement behaviour of individuals whose movements are a behavioural response to external covariates (such as ocean currents or habitat types) as well as internal motivations (such as a desire to move towards a spawning or feeding ground).

Movement behaviour is introduced using advection fields that represent preferred movement directions. The preferred direction varies with location and could also vary with time in order to reflect changes in the animal's environment or its internal state. In this way, both learning and seasonal behaviours can be modelled. More than one motivation for movement can be introduced into the same model, by using more than one advection field. The overall movement direction is given by a weighted sum of the advection fields. The coefficients that weight the advection fields are estimated parameters of the movement model.

Movement is assumed to occur on a Euclidean plane, rather than a globe. For any but the most widely migratory animals, which move from the poles to the equator and beyond, this will be a reasonable assumption. Applications to animals that are moving close to the poles would require a remapping of the spherical co-ordinates of latitude and longitude so that the pole is, effectively, at a more distant point on the globe (e.g. Bruce *et al.* , 2001).

The advection fields are calculated from co-models, each of which represents a hypothesis regarding how an animal may choose to move. If the hypothesis involves a covariate of movement, such as sea surface temperature or bathymetry, then these data are used within the co-model to calculate an advection field representing the way in which the animal responds to the covariate. In the examples presented in this thesis, the co-models are run prior to the main movement model, so that the resulting advection fields are static with time and their parameters are not estimated. However, it is theoretically possible to incorporate the co-models into the movement model so that their parameters become parameters of the movement model. This would significantly add to computational load and confounding of parameters is possible, so that exploration of this possibility is left for future work.

The modelling framework presented here has been developed for application to tracking data and can be used to compare competing hypotheses regarding movement behaviour (and therefore likely covariates of movement). Such model comparison is shown in Chapter 5.

As motivated in Chapter 2, a Bayesian state space modelling framework is considered to be highly suitable for this problem. It considers both observation and process errors, accounts for the autocorrelation inherent in track data, and is flexible enough to incorporate a wide range of behaviours. Methods are available that allow inference and model comparison for state space models (SSMs).

In the framework presented here, the usual assumption of regular time steps has been adapted so that the irregularly timed observations typical of tracking data can be modelled. Unobserved locations can be included in the model, obviating the need to make the usual assumption that a tracked animal moved in straight lines between all observed locations.

In many cases, movement behaviour will be complex, or, at least, not yet well understood. Therefore, proposed movement behaviours are unlikely to exactly resemble the real behavioural mechanism used by the tracked animal. However, as shown in

Chapter 5, model comparison methods can be used, together with the movement modelling framework presented here, to discriminate between broadly different behavioural motivations. In addition, estimation of the parameters of the movement model (without model comparison) can reveal the relative influence on the tracked individual's movements of various covariates of behaviour. If several advection fields are included in the same model, their relative importance, or weight, is given by the estimated coefficient for each. An estimated value close to zero indicates that an advection field (and the covariate, if any, that it uses) is not influencing the movement of the tracked individual.

In order to better understand this modelling tool and its ability to correctly identify the relative importance to movement of alternative advection fields (or motivators for movement), this chapter uses simulation testing to examine the model's ability to recover the parameter values used to generate simulated data, and explores bias and confounding of the parameters.

The aim of this chapter is to present a flexible modelling framework for application to irregularly timed tracking data that allows investigators to incorporate a number of quite different hypothesized behavioural responses to covariates and to internal states, using advection fields. This model is used, together with tracking data, to make inference about movement behaviours. The term 'track' is reserved for the sequence of observations of location whereas 'path' refers to the corresponding true, unobserved, locations of the animal.

3.3 Methods

A flexible state space framework is presented for modelling a wide range of hypothesized movement behaviours. This model allows for a range of choices for the residual error structures, the way in which time is discretized, and the form of the structural time series used for the state equation. From the resulting family of models one is selected for illustration.

Movement behaviour is represented in this model as a set of advection fields that direct the individual's movements. Examples are discussed in Section 3.3.3. The location of the animal determines the advection forces to which it is subjected. Each advection field is calculated prior to the implementation of the movement model and represents an idea of how the animal might be behaving. Hypotheses of movement

behaviour, and covariate information such as remotely sensed sea surface temperatures or oceanographic model-generated current data, are used to generate these advection fields (using co-models).

This example of a model drawn from the family of models represented by the framework presented here, has been implemented in R (R Development Core Team, 2007). Although WinBUGs (Spiegelhalter *et al.* , 2004) has been a successful tool for model development and inference for similar applications (Jonsen *et al.* , 2003, 2005, 2006, 2007; Morales *et al.* , 2004) it is as yet unable to facilitate the large datasets of covariate information that are likely to be required by users of this model. This model requires a Metropolis-Hastings step in order to allow for nonlinearity in the path, resulting from the dependence of the animal's location on the covariates. In addition, R provides the necessary flexibility for coding the co-models that convert this covariate information into the advection forces used by the movement model. The symbols used in this chapter are listed in Appendix 10, which also gives a brief description, and their dimensionality.

3.3.1 Observation equation

Let $x_1 \ldots x_N$ denote the true positions $x_k = (x_{k1}, x_{k2})$ of the tracked individual at a set of (possibly irregularly spaced) times $t_1 \ldots t_N$. For some subset of times $\{t_k \mid k \in O\}$, the tracking process yields an observation of the individual's location $y_k = (y_{k1}, y_{k2})$. These are related to the true locations through the observation equation

$$y_k = x_k + \epsilon_k \qquad k \in O. \tag{3.1}$$

The observation errors ϵ_k are assumed to be independent and to follow a bivariate normal distribution with precision (the inverse of variance) $\tau^y = \mathrm{diag}\,(\tau_1^y, \tau_2^y)$

$$\epsilon_k \sim \mathrm{N}\left(0, (\tau^y)^{-1}\right) \quad k \in O. \tag{3.2}$$

Only n ($n \leq N$) locations are observed, the remainder are latent.

3.3.2 State equation

The state equations describe the true (unobserved) locations $x_k = (x_{1k}, x_{2k})$ of the tracked individual's path. The difference in successive locations is modelled in terms of an inertial effect given by velocity $v_k = (v_{1k}, v_{2k})$, an advection effect $X_k\,\beta$, and an error term ξ_k. The influence of velocity and advection scales with the duration of the step $\Delta_k = t_{k+1} - t_k$

$$x_{k+1} = x_k + \Delta_k \left[v_k + (X_k\,\beta)^T\right] + \xi_k. \tag{3.3}$$

The advection terms represent the individual's tendency to move in response to both environmental factors and to internal states. Here $X_k = X_k(x_k)$ is a $2 \times m$ matrix representing the effect of m advection forces operating at x_k. The dependence on x_k is omitted for clarity, and because X_k will be independent of x_k when advection is constant (the same at all locations). The symbol β represents a vector of m regression coefficients representing the degree to which each advection force influences movement.

The inertial term, velocity v_k, represents a tendency for the individual to show directional persistence. Here v_k can be loosely interpreted as a velocity (although advection also imparts a component of velocity to the path), and is modelled as a random walk

$$v_{k+1} = v_k + \zeta_k. \tag{3.4}$$

The path ξ_k and velocity ζ_k error terms are again assumed to be independent and to have a bivariate normal distribution whose precision is a diagonal matrix $\tau^x = \mathrm{diag}\,(\tau_1^x, \tau_2^x)$, and $\tau^v = \mathrm{diag}\,(\tau_1^v, \tau_2^v)$, but now the variance scales with time step Δ_k - consistent with Brownian motion and, consequently, the process of diffusion where variance scales with time (Chakravarti, 2004),

$$\xi_k \sim \mathrm{N}\left(0, \Delta_k\,(\tau^x)^{-1}\right) \qquad k = 1 \ldots (N-1) \tag{3.5}$$

$$\zeta_k \sim \mathrm{N}\left(0, \Delta_k\,(\tau^v)^{-1}\right) \qquad k = 1 \ldots (N-2). \tag{3.6}$$

3.3.3 Modelling behaviour using advection

The advection components of the model represent the behaviour of the moving individual in response to stimuli. Their calculation may require the development of a behavioural co-model for each advection component, each representing a hypothesis regarding how the individual's behaviour is affected by that stimulus.

In the simplest scenario, a tendency to move northwards is given by a northwards advection of constant magnitude at all locations.

Somewhat more complex, a tendency to drift along with a current is given by an advection field that, at any point, has magnitude and direction equal to that of the ocean current, as determined by an oceanographic model or by observations. In this example, both magnitude and direction vary with location. The advection field, like location, is considered to be continuous in space, although it may be interpolated from a grid.

An even more complex example is given by navigation towards a specific (goal) location using the earth's magnetic field as a guide. Here a hypothesis is required regarding how the animal experiences the geomagnetic field and how it processes that information to yield a heading towards the goal location. This would be coupled with a model, or observations, of the geomagnetic field that provides a map of the geomagnetic gradients. The result would be an advection field (map) giving the direction in which the animal would move, from any given location. The geomagnetic field is not regular therefore the advection map is unlikely to give direct bearings towards the goal location. If the moving individual remembers the irregularities of the magnetic field through which it is moving, and consequently updates its mental map, then the advection field must update at each time step. If such a memory component is modelled then the advection field at any given time step would depend not only on the location of the individual at that time but also on previous locations.

Alternative hypotheses regarding how the geomagnetic field is used by the animal would yield alternative advection fields. Each alternative field, or set of fields, would be incorporated into an alternative movement model. The model comparison methods presented in Chapter 5 could then be used to establish which model, and therefore which hypothesis, is given greatest support by the observed track.

Each advection component in this model is thus given by a co-model, which combines either environmental data or suggested internal state with a theory of behaviour to

yield an advection field giving a direction and magnitude for all locations that a moving animal might occupy.

3.3.4 Posterior, Likelihood, Priors and Initial conditions

Schnute (1994) presents a general framework for state space models, allowing for nonlinear state equations and non-Gaussian error. His formulation encompasses the Bayesian context in which the complete parameter set includes both parameters θ and unknown states (in this case x and v). In a Bayesian analysis there is no formal distinction between state variables and parameters, all are treated as random variables (de Valpine, 2002). Nevertheless, for clarity, in this work the terms 'parameter' and 'state variable' are used separately.

In Schnute's formulation, V is a control vector that represents observed data that, unlike y (and X), do not depend directly on x. It is assumed to be known without error. The posterior $p(x, v, \theta \mid y, V)$ is proportional to

$$p(x, v, \theta \mid y, V) \propto p(y \mid x, v, \theta, V)\, p(x \mid v, \theta, V)\, p(v \mid \theta, V)\, p(\theta). \tag{3.7}$$

Here $p(x \mid v, \theta, V)$ is termed the prior for x and similarly $p(v \mid \theta, V)$ is the prior for v.

The advection term X may be dependent on x, making this problem highly nonlinear. In this formulation V represents the totality of environmental and behavioural information to which the individual might respond. The advection term X represents the subset of V to which the individual actually responds. For example, in an oceanographic context, V might represent the entire remotely sensed sea surface temperature (SST) field, in which case X would represent the SST values at the individual's true locations x. In the equations below, the symbol V is used to indicate dependence on control data (advection), but X is used to indicate that the specific values $X(x)$ are required.

The probability of the data y given the full parameter set is $p(y \mid x, v, \theta, V)$. This density is dependent on only the true locations x and the precision parameter for the observation errors, through equation 3.2. Because the observation errors are assumed to be independent between time steps, this density is a product of n bivariate normals (equation 3.2)

$$p(y \mid x, v, \theta, V) = \prod_{k \in O} p(y_k \mid x_k, \theta). \tag{3.8}$$

In fact, because the covariance terms are zero (equation 3.2) this is a product of $2n$ normals.

The density for the true locations given the velocities $p(x \mid v, \theta, V)$ is the product of the prior placed on the first location, and the location errors (from equation 3.5). It is assumed that these errors are independent between time steps, again yielding the product of bivariate normals (and in fact, again, of $2(N-1)$ normals)

$$p(x \mid v, \theta, V) = p(x_1 \mid \theta) \prod_{k=2}^{N} p(x_k \mid x_{k-1},\ v_{k-1}, \theta,\ X_{k-1}). \tag{3.9}$$

Similarly, the density for the velocity terms given the parameters and control vector V is the product of a prior on the first velocity v_1 (an initial condition) and the errors from the stochastic process for velocity (equation 3.6)

$$p(v \mid \theta, V) = P(v_1 \mid \theta) \prod_{k=2}^{N-1} p(v_k \mid v_{k-1}, \theta). \tag{3.10}$$

To reflect uncertainty in knowledge of the first location and first velocity, bivariate normal priors are placed on these, again the variance is a diagonal matrix $\tau^{x0} = \text{diag}\left(\tau_1^{x0},\ \tau_2^{x0}\right)$. These represent the model's initial conditions

$$x_1 \sim \text{MVN}_2\left(x_0,\ \left(\tau^{x0}\right)^{-1}\right) \tag{3.11}$$

$$v_1 \sim \text{MVN}_2\left(v_0,\ \left(\tau^{v0}\right)^{-1}\right) \tag{3.12}$$

These priors are selected for conjugacy with the normal likelihood. A multivariate normal prior is assumed for β

$$\beta \sim \mathrm{MVN_m}\left(\beta_0,\ \Sigma_0\right) \tag{3.13}$$

and conjugate gamma priors are assumed for the precisions for the observations

$$\tau_i^y \sim \mathrm{Gamma}\left(a_i^y,\ b_i^y\right), \quad i = 1, 2 \tag{3.14}$$

and similarly for the precisions for the path locations τ_i^x and velocities τ_i^v.

3.3.5 Conditional distributions

Having used conjugate priors, it is possible to calculate the exact conditional distributions for the parameters (τ^x, τ^v, τ^y, and β) and the v states but not for the x states in cases where X depends on x. The method for calculating these conditional distributions is described in Appendix 11. Specific calculations for a gamma, and for a multivariate normal conditional are given in Appendix 12 and Appendix 13. These pertain to the model presented in Chapter 4 but are very similar to those presented here, except that they present some additional complications because in that chapter β is allowed to vary with time.

Conditional distributions for all parameters and states are presented below, but first the three error terms are expressed as matrices.

The observation errors ϵ (equation 3.1) are functions of the observed locations y and their corresponding true locations $P\,x$

$$\epsilon = y - P\,x. \tag{3.15}$$

Here P is an $n \times N$ matrix of zeros and ones that drops the locations for which no observation has been made, while retaining ordering, leaving only the subset of x relating to the time steps $k \in O$ during which an observation is made

$$P_{jk} = \begin{cases} 1 & \text{if the } j\text{th observation is made during time step } k \\ 0 & \text{otherwise.} \end{cases} \tag{3.16}$$

The location errors ξ (equation 3.3) are functions of the difference in the true locations $D_x x$ (D_x is a matrix that produces the $N-1$ first differences of x), the time steps $\Lambda_x = \text{diag}(\Delta)$, the velocities v, and the partitioned advection matrix $[\mathcal{X}_1 \beta \mid \mathcal{X}_2 \beta]$

$$\xi = D_x x - \Lambda_x(v + [\mathcal{X}_1 \beta \mid \mathcal{X}_2 \beta]). \tag{3.17}$$

Here $\mathcal{X}_1$ and $\mathcal{X}_2$ are $N-1 \times m$ matrices such that the kth row of $\mathcal{X}_i$ is the ith row of X_k.

Velocity (equation 3.4) is a random walk so that the errors in velocity ζ are simply the first differences of v

$$\zeta = D_v v \tag{3.18}$$

where D_v is the matrix that produces the $N-2$ first differences of v.

When stating the conditional distributions for the parameters, $\theta(-\tau_i^y)$ is used to indicate all parameters except τ_i^y. The conditionals for the precision terms are gamma distributions. The conditional for the precision for the observations τ_i^y, $i = 1 \ldots 2$ is

$$\tau_i^y \mid \theta(-\tau_i^y), x, v, y, V \sim \text{Gamma}\left(\frac{N}{2} + a_i^y,\ \frac{1}{2}\,\epsilon_i^T\,\epsilon_i + b_i^y\right), \tag{3.19}$$

and for the precision for the path τ_i^x, $i = 1 \ldots 2$ is

$$\tau_i^x \mid \theta(-\tau_i^x), x, v, y, V \sim \text{Gamma}\left(\frac{N-1}{2} + a_i^x,\ \frac{1}{2}\,\xi_i^T\,\Lambda_x^{-1}\xi_i + b_i^x\right), \tag{3.20}$$

and for the precision for the velocities τ_i^v, $i = 1 \ldots 2$ is

$$\tau_i^v \mid \theta(-\tau_i^v), x, v, y, V \sim \text{Gamma}\left(\frac{N-2}{2} + a_i^v,\ \frac{1}{2}\,\zeta_i^T\,\Lambda_v^{-1}\,\zeta_i + b_i^y\right). \tag{3.21}$$

The conditional distribution for the advection coefficients β is a multivariate normal

$$\beta \mid \theta(-\beta), x, v, y, V \sim \text{MVN}_\text{m}\left(\mu^\beta,\ \Sigma^\beta\right) \tag{3.22}$$

with mean

$$\mu^\beta = \Sigma^\beta \sum_{i=1}^{2} \left[\tau_i^x\ \mathcal{X}_i^T\ (D_x\,x_i - \Lambda_x v_i)\right] + \Sigma_0^{-1}\,\beta_0$$

and variance

$$\Sigma^\beta = \left[\Sigma_0^{-1} + \sum_{i=1}^{2} \tau_i^x\ \mathcal{X}_i^T\ \Lambda_x\ \mathcal{X}_i\right]^{-1}.$$

The conditional distribution for the inertial state (velocity) v is multivariate normal

$$v_i \mid \theta, x, v_{-i}, y, V \sim \text{MVN}_{\text{N-1}}\left(\mu_i^v, \Sigma_i^v\right) \tag{3.23}$$

with mean

$$\mu_i^v = \Sigma_i^v\ \left[\tau_i^x\ (D_x\,x_i - \Lambda_x\,\mathcal{X}_i\,\beta) + \tau_i^{v0}\ (v_0)_i\,\mathcal{I}_{v0}\right]$$

and variance

$$\Sigma_i^v = \left[\tau_i^x \, \Lambda_x + \tau_i^v \, D_v^T \, \Lambda_v^{-1} \, D_v + \tau_i^{v0} \, \mathcal{I}_{v0} \, \mathcal{I}_{v0}^T\right]^{-1}.$$

Here, Λ_v is a diagonal matrix of time steps excluding the last Δ_{N-1}, and $\mathcal{I}_{v0}$ is a column vector of length $N-1$ of the form $(1,0,0,...,0)^T$ that increases the dimension of v_0 and τ_i^{v0} by adding zeros.

If advection were constant ($\mathcal{X}$ independent of x) the conditional for x could be calculated, and would be a multivariate normal. The symbol $x_i \mid x_{-i}$ is used to indicate the vector of longitudes for all locations given the vector of latitudes for all locations (or all latitudes given all longitudes)

$$x_i \mid \theta, x_{-i}, v, y, V \sim \mathrm{N}\left(\mu_i^x, \, \Sigma_i^x\right) \tag{3.24}$$

with mean

$$\mu_i^x = \Sigma_i^x \, \left[\tau_i^y \, P^T \, y_i + \tau_i^x \, D_x^T \, (\mathcal{X}_i \, \beta + \, v_i) + \tau_i^{x0} \, \mathcal{I}_{x0} \, (x_0)_i\right]$$

and variance

$$\Sigma_i^x = \left[\tau_i^y \, P^T \, P + \tau_i^x \, D_x^T \, \Lambda_x^{-1} \, D_x + \tau_i^{x0} \, \mathcal{I}_{x0} \, \mathcal{I}_{x0}^T\right]^{-1}$$

where $\mathcal{I}_{x0}$ is a column vector of length N of the form $(1,0,0,...,0)^T$ that increases the dimensionality of x_0 and τ_i^{x0} by adding zeros.

3.3.6 MCMC

Conditional distributions have been calculated for the parameters (τ^x, τ^v, τ^y, and β) and for the v states. Had a conditional for the x states been available, Gibbs sampling could have been used to make draws from the joint posterior distribution and thereby the marginal distributions for all parameters and states. Because a Bayesian form has been used, all parameters and states are considered to be random variables. For brevity, the term 'parameter' is used in the following brief description of Gibbs

sampling to mean both parameters and states.

Given a set of starting values for all but one parameter, Gibbs sampling proceeds by making a random draw from the conditional distribution for that parameter. The new value joins the set of starting values and a draw is made from the conditional distribution for another parameter. This proceeds until one draw has been made from the conditionals for all parameters. These actions represent a single iteration of the sampler. The resulting set of values represents one draw from the posterior distribution, and becomes the starting set for the next iteration. Each set of values is stored, and sampling proceeds at least until the investigator is satisfied that the sample has converged on the true posterior distribution.

Measurement of convergence is difficult because the underlying distribution is unknown and the sample is inherently autocorrelated. Nevertheless, a number of indicators of convergence have been presented, although their authors acknowledge that they can fail to accurately diagnose convergence (Gelman & Rubin, 1992; Geweke, 1992; Roberts, 1992; Cowles & Carlin, 1996). These have been incorporated into the CODA software, which is accessible through the R programming language (R Development Core Team, 2007) and has been integrated into the WinBUGs package (Spiegelhalter *et al.* , 2004).

Lacking a conditional distribution for x, this component of the Gibbs iteration is replaced with a Metropolis-Hastings step. Let b represent a proposed set of values for x that is drawn from a proposal distribution $q(\cdot, \cdot)$, which is selected for its similarity to the true conditional $p(x \mid \theta, v, y, V)$. This proposal is accepted, the set of values for x from the previous iteration a, with probability $\alpha(a, b)$

$$\alpha(a, b) = \min\left\{1, \frac{\pi(b)\, q(b, a)}{\pi(a)\, q(a, b)}\right\}. \tag{3.25}$$

where $\pi(\cdot)$ is the posterior distribution. This is known up to an unknown normalizing constant but because $\pi(\cdot)$ appears in both the numerator and denominator this constant is cancelled out. The proposal distribution $q(a, b)$ represents a step from a to b. The 'conditional' for x (equation 3.24), which is a true conditional only if the whole advection field V is constant at all locations, was used as the proposal distribution

$$q(a, b) = p\left(b \mid \theta, v, y, X(a)\right) \tag{3.26}$$

and similarly for $q(b, a)$. Note that the notation used here differs from equation 3.24 in that, for clarity, the subscript i denoting Cartesian coordinate, has been dropped, along with the reference to x_{-i}. This indicates that x can be updated in one step because x_1 and x_2 are independent of one another.

If advection takes the same values at all locations then $X(a) = X(b) = X$ and $q(\cdot, \cdot)$ is the true conditional for x, which is equal to the posterior multiplied by a term c that is constant with respect to x. The values for all parameters and the v states are held constant while performing this Metropolis-Hastings step for x so that, if advection is constant

$$q(a, b) = p\left(b \mid \theta, v, y, X\right) = \pi(b)\, c \tag{3.27}$$

And similarly for $q(b, a)$. Substituting equation 3.27 into equation 3.25, the acceptance probability for the Metropolis-Hastings step (when advection is constant) becomes

$$\alpha(a, b) = \min\left\{1, \frac{\pi(b)\, \pi(a)\, c}{\pi(a)\, \pi(b)\, c}\right\} = 1 \tag{3.28}$$

so that this step reverts to Gibbs sampling when advection is constant.

The first 1000 MCMC draws were discarded in order to reduce the effect of the starting values. The subsequent 200 000 iterations were retained. Convergence was checked by running several chains, each with different starting points and ensuring that the results did not differ noticeably among chains. The CODA diagnostic software was used to test for convergence of the chains.

3.3.7 Simulation scenarios

This model can be used in two ways in order to make inferences about the movement behaviour of animals, given observed tracks. In both cases, hypotheses regarding

how animals might be moving must first be formulated (possibly using covariate data), and these must be used to generate advection fields (which then represent those hypotheses). First, the movement model, with more than one advection field, can be applied to track data and the coefficients of the advection fields estimated. Examination of the marginal posterior distributions for the advection coefficients will indicate whether or not the animal is responding to each advection field, and what relative importance each field has for movement.

Second, more than one version of the movement model can be produced, each using a different (set of) advection field(s). In this case the estimated parameter values of the models are not of interest and model comparison methods (explored in Chapter 5) are used to measure the weight given by the observed track to each competing model, thus sifting more likely hypotheses from less likely ones. In this case, the whole movement model represents the hypothesis regarding movement behaviour.

The first of these scenarios is explored in this chapter, and the second in Chapter 5. The present chapter uses simulation tests to explore the model's ability to correctly estimate parameter values. In order to do this, a particular form of model was selected from the family of models made available by the modelling framework presented above. This was used, together with the advection fields described below, to generate simulated track data. The model was then fitted to this data and the resulting parameter estimates were compared with their true values.

Four advection scenarios are considered (Figure 3.1). First, the simplest kind of advection field is used - one that represents constant movement towards the north (Figure 3.1a); here advection and inertia are likely to operate in the same direction and, when this advection field is used for estimation, the path x is estimated by Gibbs sampling. Second, a gyre is used to represent movement around a central point. Here the direction of advection changes with every step so it is likely to operate in a different direction from inertia (Figure 3.1b). Next, in order to test the model's ability to discriminate the coefficients of two advection fields, combinations of advection forces were specified. The third simulation scenario uses two constant advection fields, a northerly flow combined with an easterly flow, giving an overall movement to the northeast (Figure 3.1c). Again, when used in estimation, this advection scenario would result in Gibbs sampling for the path x. The fourth advection scenario combines a constant northerly flow with a clockwise gyre (Figure 3.1d), requiring Metropolis-Hastings within Gibbs sampling because of the nonlinearity introduced by the gyre.

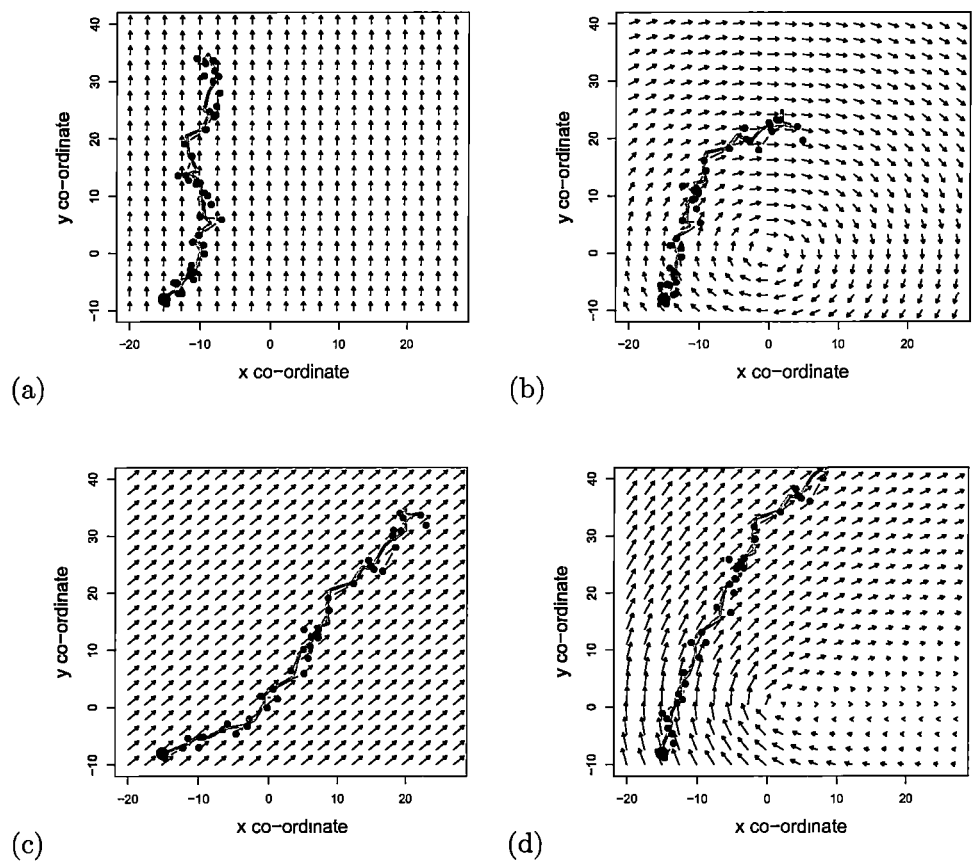


Figure 3.1: Advection fields, or combinations of fields, for the four simulations used to generate tracks; (a) a single northwards flow; (b) a single clockwise gyre; (c) combined northwards and eastwards flows; and (d) combined northwards flow with clockwise gyre. Plots (c) and (d) show only the vector sum of the two fields that are used in these scenarios. The path x that was simulated using each (set) of fields is shown in grey, with a large black dot indicating the first location. The observed locations (track) y are shown as black, joined, dots.

The four simulation scenarios were used to generate four tracks for individuals moving according to each (set of) advection field(s) (Figure 3.1). The parameter values used for these simulations are shown in Table 3.1. In order to allow the observed y and true x locations to vary to, what is hoped to be, a realistic degree, precision terms of $\tau^x = \tau^y = (1, 1)$ were used (variance, correspondingly is = 1). However, velocity v is used to introduce inertia into the path; an imprecise τ^v will allow frequent, sudden, direction changes whereas a precise τ^v will force the track to be smoother. Note that an investigator could choose to exclude errors from the velocity state v altogether. As there is no reason to allow this error term to be large, a relatively precise value was chosen for the simulation $\tau^v = (100, 100)$.

A relatively small number of observed locations (=21) is assumed, in order to simulate a data poor situation. Real applications of this model are likely to involve a greater number of observations, but these may be offset by greater variability in advection fields that are calculated using environmental data. A latent location is inserted between every pair of observations, giving a total of 41 locations to estimate.

Unequal time steps were generated from a normal distribution.

Table 3.1: 'True' parameter values used to simulate tracks, are shown. Note that the shape a and rate b parameters of the gamma distribution can be calculated from the distribution's mean μ and variance σ^2 using the formulae $a = \mu^2/\sigma^2$ and $b = \mu/\sigma^2$. The gamma prior distributions are expressed both in terms of shape a and rate b, and mean and variance.

Symbol	Description	Value
N	Number of locations	41
n	Number of observed locations	21
m	Number of advection fields	1 or 2
x_1	Initial location	(−15, −8)
v_1	Initial velocity	(0, 0)
Λ	Step durations (drawn randomly)	$\sim$ N(0.75, 0.3)
β	Coefficient of advection	1 or (1, 1)
O	Index of observed locations	$1, 3, \ldots 41$
τ^y	Precision for observed locations y	(1, 1)
τ^x	Precision for locations x	(1, 1)
τ^v	Precision for velocity v	$(0.01^{-1}, 0.01^{-1})$

In all cases, the uncertainty likely to prevail regarding the true values of all of the precision terms (the path τ^x, the observations τ^y, and the velocities τ^v) was reflected by using priors whose variance was (1, 1), (Table 3.2).

No initial direction, or magnitude, was assumed for velocity $v_0 = (0,\ 0)$, and this prior was chosen to be precise, $\tau^{v0} = (1000,\ 1000)$. The first location on the track is likely to be a release location and therefore known with precision. Its prior was centred on the correct value, $x_0 = (-15,\ -8)$, and was precise, $\tau^{x0} = (0.001^{-1},\ 0.001^{-1})$, Table 3.2.

In all estimation scenarios, the estimation model used a prior for the advection coefficients centred on zero $\beta_0 = 0$, favouring little or no influence on movement by any advection field. Therefore, influence will not be assumed unless the data indicate otherwise.

Table 3.2: Priors used for estimation. Note that the shape a and rate b parameters of the gamma distribution can be calculated from the distribution's mean μ and variance σ^2 using the formulae $a = \mu^2/\sigma^2$ and $b = \mu/\sigma^2$. The gamma prior distributions are expressed both in terms of shape a and rate b, and also mean and variance.

Symbol	Description	Value
x_0	Mean initial location	$(-15,\ -8)$
v_0	Mean initial velocity	$(0,\ 0)$
τ^{x0}	Precision for initial location	$(1000,\ 1000)$
τ^{v0}	Precision for initial velocity	$(1000,\ 1000)$
β_0	Mean of multivariate normal prior for β	0.0
Σ_0	Variance for prior for β	1
a^y	Shape for gamma prior for τ^y	(1, 1) or (25, 25)
b^y	Rate for gamma prior for τ^y	(1, 1) or (5, 5)
	Mean for gamma prior for τ^y	(1, 1) or (5, 5)
	Variance for gamma prior for τ^y	(1, 1)
a^x	Shape for gamma prior for τ^x	(1, 1) or (25, 25)
b^x	Rate for gamma prior for τ^x	(1, 1) or (5, 5)
	Mean for gamma prior for τ^x	(1, 1) or (5, 5)
	Variance for gamma prior for τ^x	(1, 1)
a^v	Parameters for gamma prior for τ^v	$(10^4,\ 10^4)$ or $(25e^4,\ 25e^4)$
b^v	Parameters for gamma prior for τ^v	(100, 100) or (500, 500)
	Mean for gamma prior for τ^v	(1, 1) or (500, 500)
	Variance for gamma prior for τ^v	(1, 1)

Three investigations were carried out. The first examined the ability of the movement model to correctly estimate parameter and state values under ideal conditions. The same advection field(s) used to simulate the track was used in the estimation model, and the priors for the precision terms were centred on their simulation values. Four applications occurred, one for each of the advection scenarios shown in Fig-

ure 3.1. Two of these scenarios involved two advection fields, used in combination, so that the estimator's ability to discriminate their coefficients could be examined (Figure 3.1c and d). Two scenarios used constant advection fields so that the Gibbs sampler was used for all parameters and states (Figure 3.1a and c) whereas the remainder required a Metropolis-Hastings step for the path x (Figure 3.1b and d).

The second investigation examines sensitivity to the prior for the precision parameters. This was examined by using a prior whose mean was five times that of the value used to simulate the data (Table 3.2); here, the track generated using the simplest advection scenario (constant northerly flow), Figure 3.1a) was used in the estimation model.

The third investigation examines sensitivity to incorrect advection fields. This was examined by applying an estimation model that assumed a constant northerly flow (Figure 3.1a) to a track that was simulated using the most complex advection scenario, northerly flow combined with a gyre (Figure 3.1d).

3.4 Results

When a single advection field is used in both the simulation and estimation of the track, and the priors for the precision parameters are centred on the values used in the simulation, all parameters and states are estimated without apparent bias (Figures 3.2 and 3.3). A possible exception is the precision for the observation errors τ^y, which has a heavy right-hand tail; but, being gamma distributed, some skewness is expected. The scenario that uses a constant northerly advection (Figure 3.2) has an oblong cloud of estimates of velocity as compared with the circular cloud for the scenario that uses a gyre (Figure 3.3). The constant advection scenario has both inertia and advection operating in the same direction so that there is greater uncertainly regarding the value of velocity in the north-south direction compared with the east-west direction. This confusion does not occur in the gyre scenario because the direction of advection changes with time (Figure 3.3). Similarly, the estimates of location for the northerly scenario (Figure 3.2) are more spread in the east-west direction than the north-south direction whereas those for the gyre are more evenly spread in all directions.

The precision for velocity τ^v is estimated with greater precision than those for observation τ^y and location τ^x because of the more precise mean value assumed for

this prior. Researchers may prefer to use this parameter, along with a precise prior, to set the smoothness of the path, rather than estimating it.

Path location is estimated with greater precision when advection acts more strongly on movement. This can be seen by comparing the location estimates x for the scenarios that use a single advection field (Figures 3.2 and 3.3) with those that use two combined advection fields (Figures 3.4 and 3.5). In all cases the advection coefficient used to simulate the track was the same, $\beta = 1$, so that when the fields operate in the same direction the combined effect is doubled, making it easier for the model to discriminate between advection and random error.

This paragraph discusses *bias* in the estimated parameter values, where an unbiased estimator is one that produces estimates whose mean or median is equal to the true parameter value (Gelman *et al.* , 2004). In the histograms referred to here, such lack of bias will be clear from a symmetrical distribution of the posterior about the true value (indicated in the by shaded bars). The estimators that used a single advection field produced unbiased estimates of the coefficients β (Figures 3.2 and 3.3), however those that use two advection fields show some bias (Figures 3.4 and 3.5). The model that uses northerly combined with easterly advection (Figure 3.4) has an unbiased estimate of the coefficient for advection in the northerly direction β_1 but that for easterly advection β_2 is overestimated, indicating stronger movement towards the east. This is balanced by a westward bias in velocity v, while there is no bias for v in the north-south direction. This indicates confounding of velocity (inertia) with the advection coefficients β. Once the animal had begun to move in an easterly direction, the tendency to continue doing so is attributed to inertia instead of to advection.

The track that was simulated using northerly advection together with a gyre moves through a region in which the gyre has a largely northerly flow (Figure 3.1d) so that the coefficients for the two advection fields are confounded. The coefficient for northerly flow is underestimated, while that for the gyre is overestimated (Figure 3.5).

Centring the prior for a precision parameter on an incorrect mean value (five times the correct value) causes the marginal for that parameter's posterior to shift correspondingly (Figures 3.6 - 3.8), indicating that the data provide little information on the magnitude of these quantities. When the prior for the observation precision τ^y (Figure 3.6) has an incorrect mean, the marginal posterior for the path precision

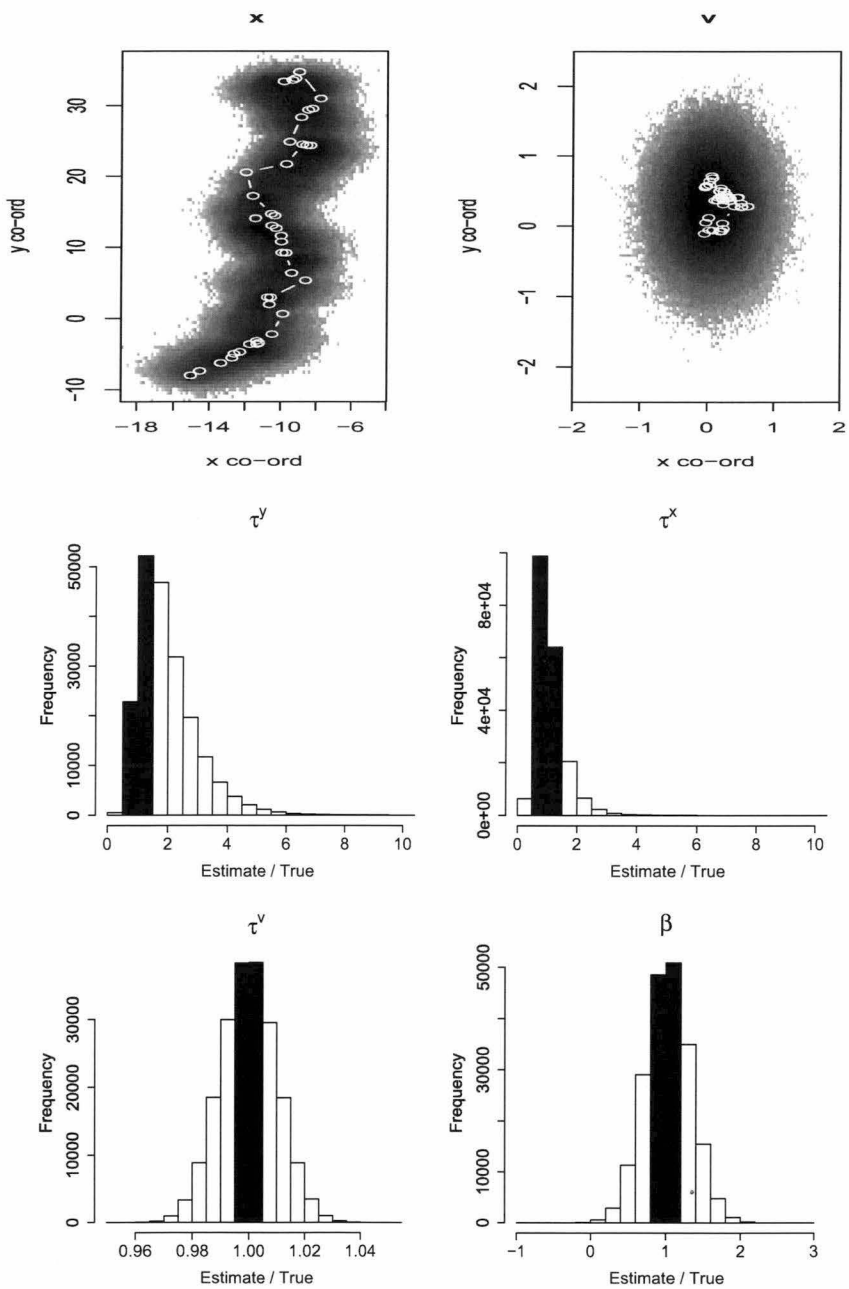


Figure 3.2: Results from 200 000 MCMC draws from the posterior of the estimation model that uses northerly advection (scenario 1), applied to the track generating using this same advection. States are shown as Cartesian coordinates (the simulated values as open circles joined by white lines and the MCMC draws as a density plot). Parameters are shown as a fraction of their true values so that a correct value is given as 1; the bars that span this value are shaded in black.

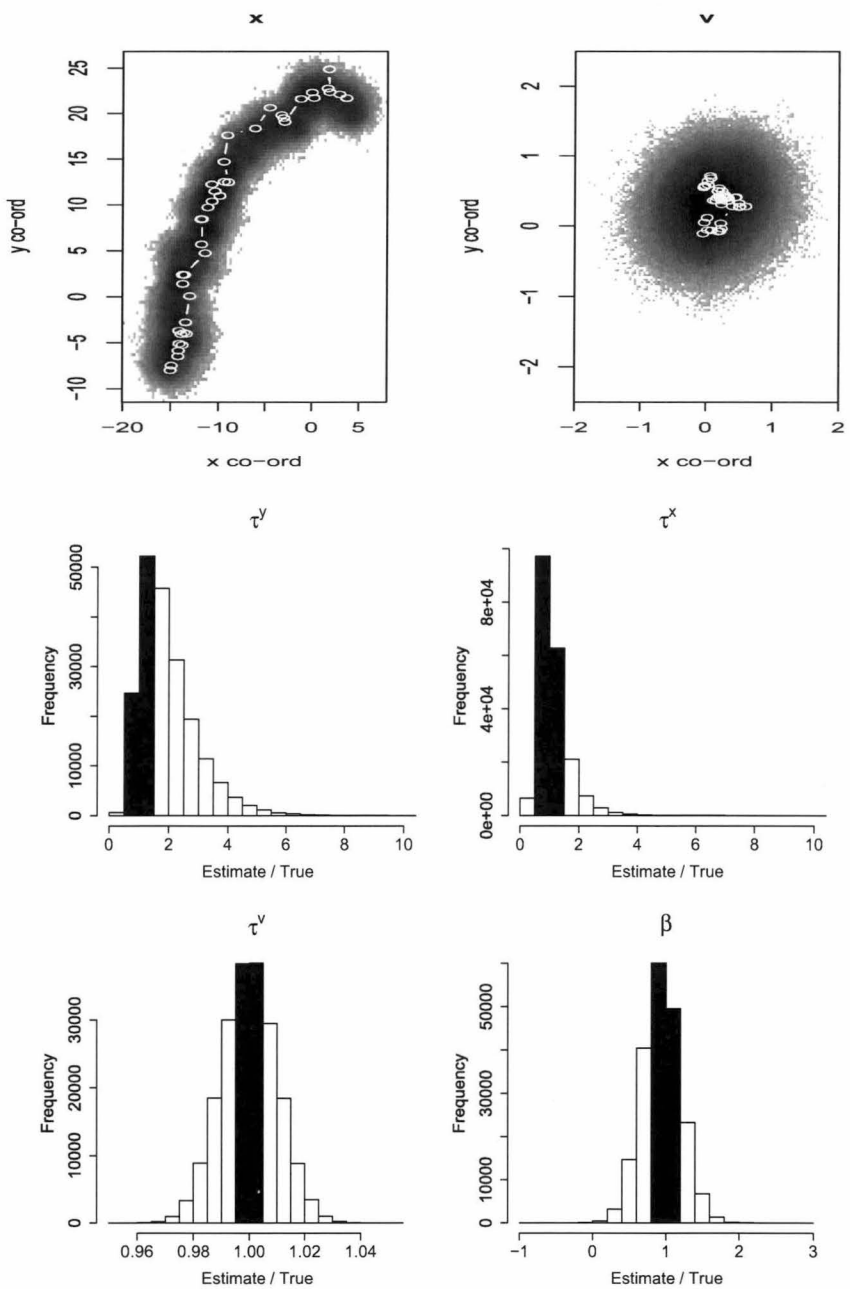


Figure 3.3: Results from 200 000 MCMC draws from the posterior of the estimation model that uses a clockwise gyre (scenario 2), applied to the track generating using this same advection. States are shown as Cartesian coordinates (the simulated values as open circles joined by white lines and the MCMC draws as a density plot). Parameters are shown as a fraction of their true values so that a correct value is given as 1; the bars that span this value are shaded in black.

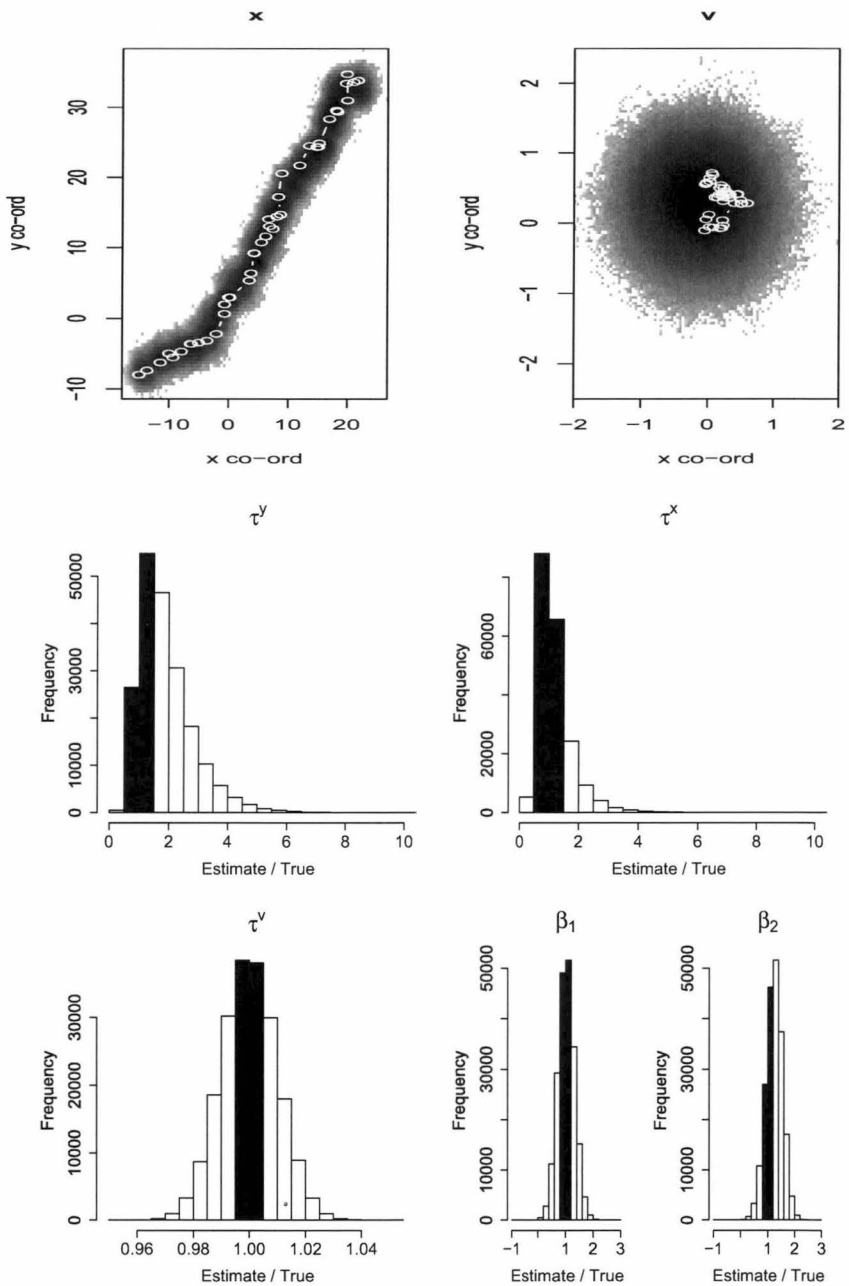


Figure 3.4: Results from 200 000 MCMC draws from the posterior of the estimation model that uses northerly advection combined with easterly advection (scenario 3), applied to the track generating using these same advections. States are shown as Cartesian coordinates (the simulated values as open circles joined by white lines and the MCMC draws as a density plot). Parameters are shown as a fraction of their true values so that a correct value is given as 1; the bars that span this value are shaded in black.

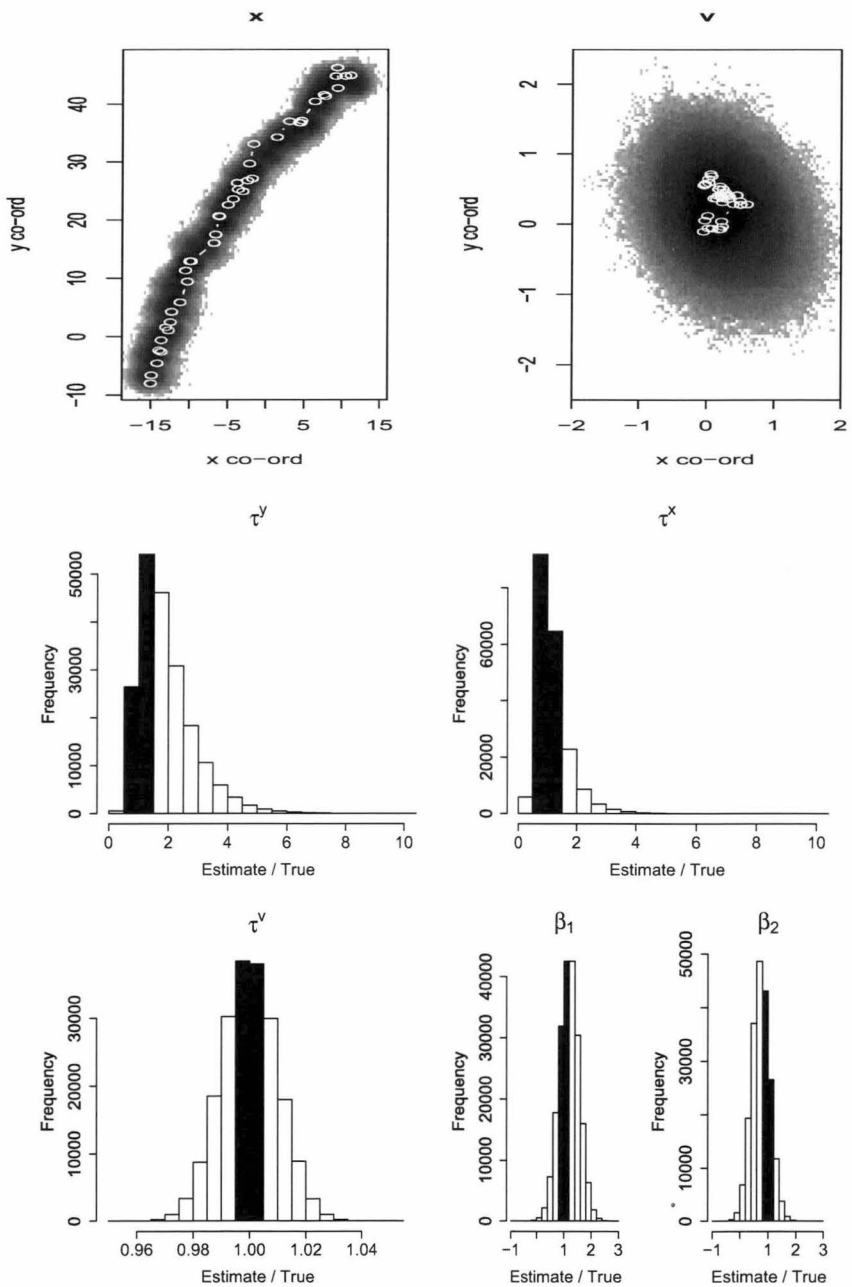


Figure 3.5: Results from 200 000 MCMC draws from the posterior of the estimation model that uses northerly advection combined with a gyre (scenario 4), applied to the track generating using these same advections. States are shown as Cartesian coordinates (the simulated values as open circles joined by white lines and the MCMC draws as a density plot). Parameters are shown as a fraction of their true values so that a correct value is given as 1; the bars that span this value are shaded in black.

τ^x becomes more precise. Similarly, when the prior for τ^x has an incorrect mean, the marginal posterior for τ^y becomes more precise (Figure 3.7), indicating that the ratio of these parameters is more precisely determined than their absolute values.

Reassuringly, misspecification of the precision terms τ^y and τ^x have little or no effect on the marginal for the advection coefficient β, except for a slight increase in precision when τ^x is overestimated (Figure 3.7).

Misspecification of the prior, and therefore a shift in the posterior, for the velocity precision τ^v has no effect on the estimates of τ^y or τ^x but causes a slight upward bias in the marginal for the advection coefficient β (contrast Figure 3.8 with Figure 3.2). This indicates that for the range of values investigated here, τ^v may be uncorrelated with the other parameters and the x states and can therefore be treated as part of the choice of model drawn from the family of possibilities presented by the modelling framework, rather than as a parameter that must be estimated. This upward shift in τ^v (indicating increased precision, decreased variance) does result in greater precision in the estimation of the velocities v.

When the northerly advection and gyre (Figure 3.1d) were used to generate the track but only the northerly flow (Figure 3.1a) was used in estimation, a bias appeared in the posterior marginal velocity - the centre shifting to the southeast (Figure 3.9). In the region in which this animal is moving, the gyre exerts a northeasterly to easterly pressure. The northerly component of this explains the upward bias in the posterior for the coefficient for northerly flow (β North in Figure 3.9) and the shift in the posterior for v compensates for the easterly movement and to some extent the reduced northerly movement in the latter part of the track.

The marginal posteriors for v for the scenarios that used correct advection fields were centred on (0,0), indicating no persistent directionality to inertia. The scenario that used an incorrect advection field shows a shift in the centre of the posterior for v to roughly (1,-0.5), Figure 3.9.

3.5 Discussion

A modelling framework is presented that is capable of estimating the parameters and states of a model applied to irregularly timed, autocorrelated tracking data. The model simulated an individual's response, in terms of movement behaviour,

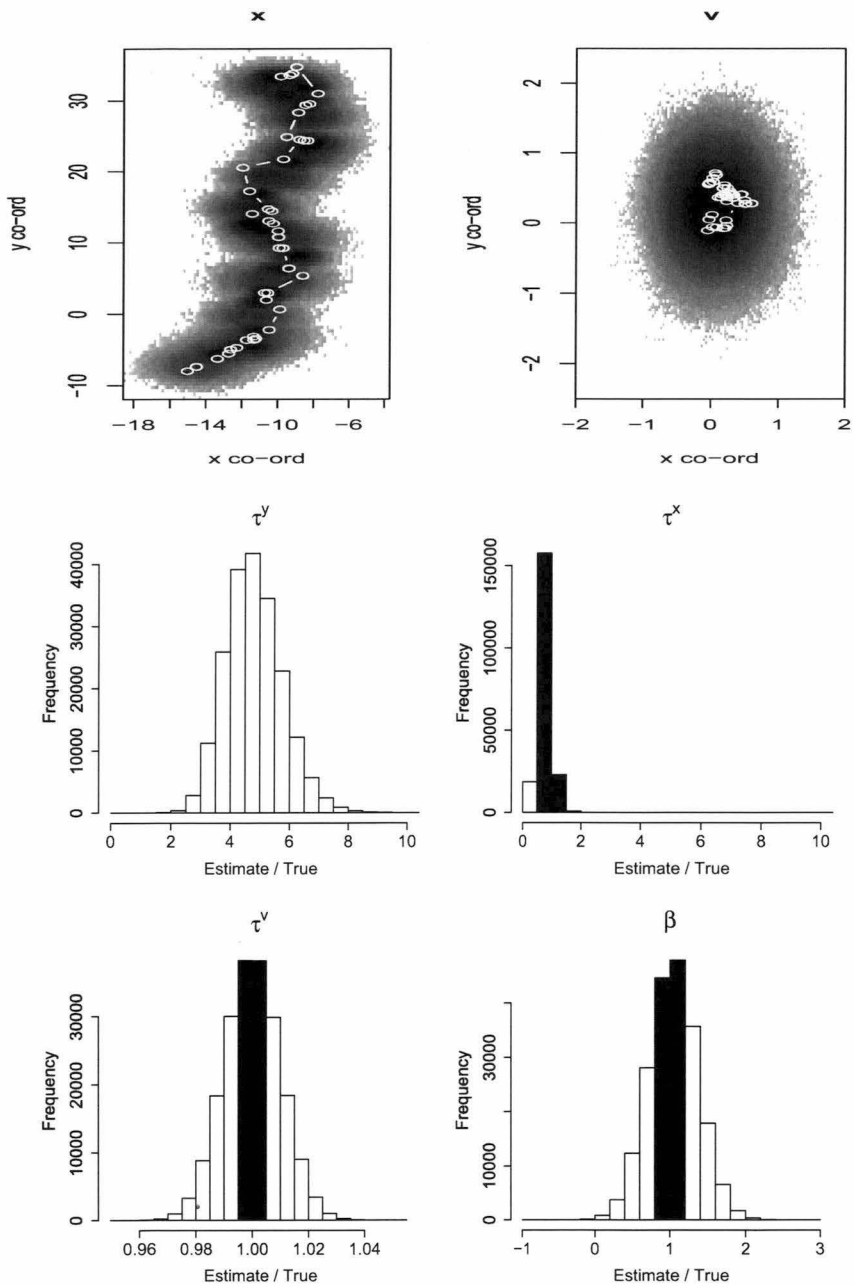


Figure 3.6: Results from 200 000 MCMC draws from the posterior of the estimation model that uses northerly advection (scenario 1), applied to the track generating using this same advection. The prior for τ^y had a mean value five times the true value. States are shown as Cartesian coordinates (the simulated values as open circles joined by white lines and the MCMC draws as a density plot). Parameters are shown as a fraction of their true values so that a correct value is given as 1; the bars that span this value are shaded in black.

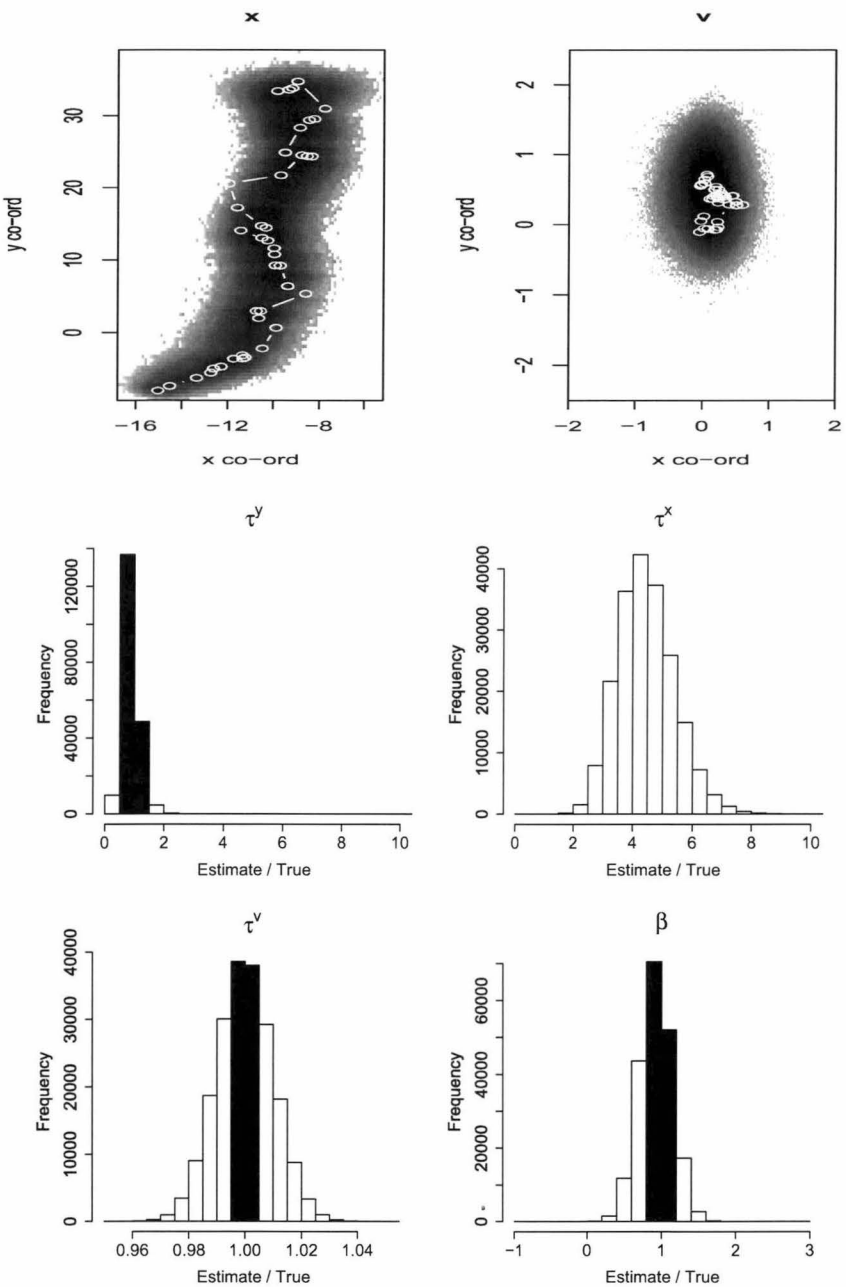


Figure 3.7: Results from 200 000 MCMC draws from the posterior of the estimation model that uses northerly advection (scenario 1), applied to the track generating using this same advection. The prior for τ^x had a mean value five times the true value. States are shown as Cartesian coordinates (the simulated values as open circles joined by white lines and the MCMC draws as a density plot). Parameters are shown as a fraction of their true values so that a correct value is given as 1; the bars that span this value are shaded in black.

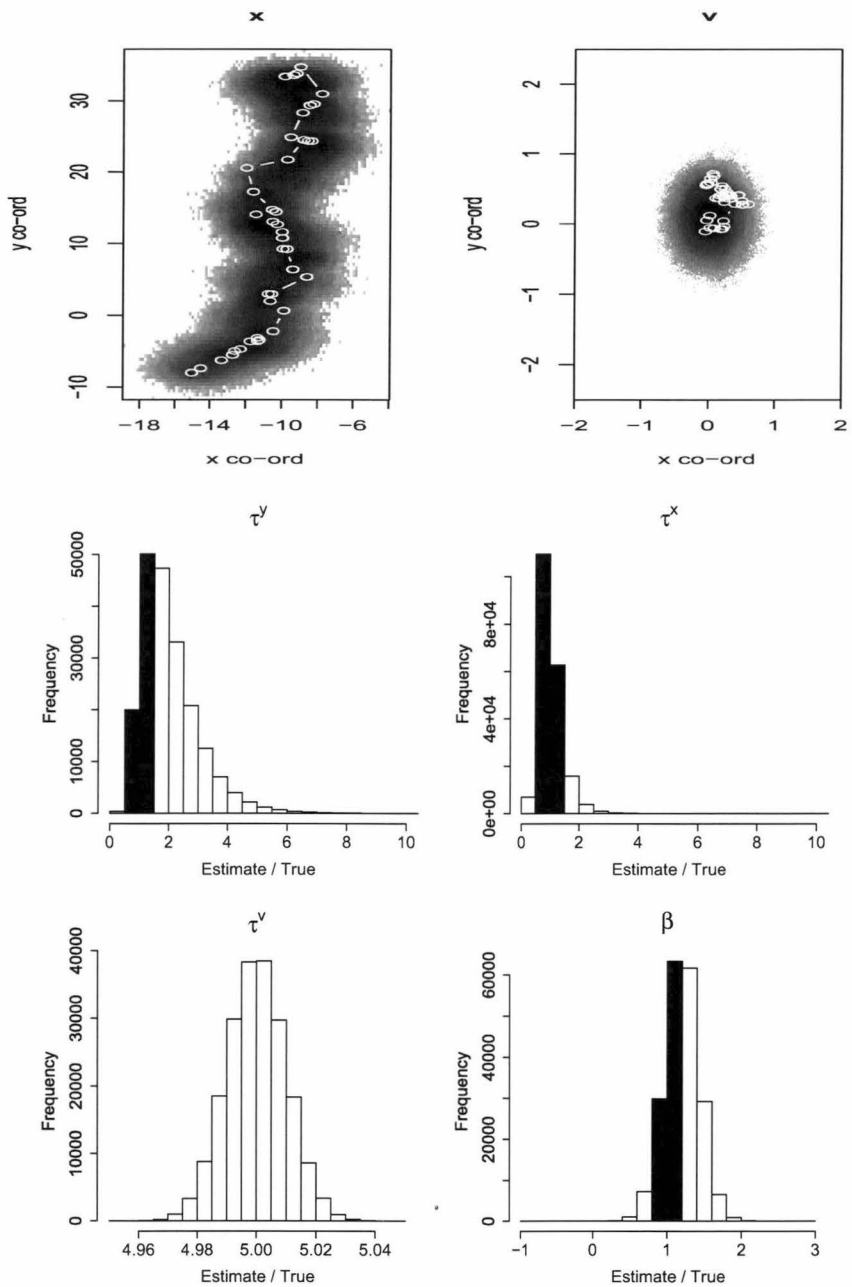


Figure 3.8: Results from 200 000 MCMC draws from the posterior of the estimation model that uses northerly advection (scenario 1), applied to the track generating using this same advection. The prior for τ^v had a mean value five times the true value. States are shown as Cartesian coordinates (the simulated values as open circles joined by white lines and the MCMC draws as a density plot). Parameters are shown as a fraction of their true values so that a correct value is given as 1; the bars that span this value are shaded in black.

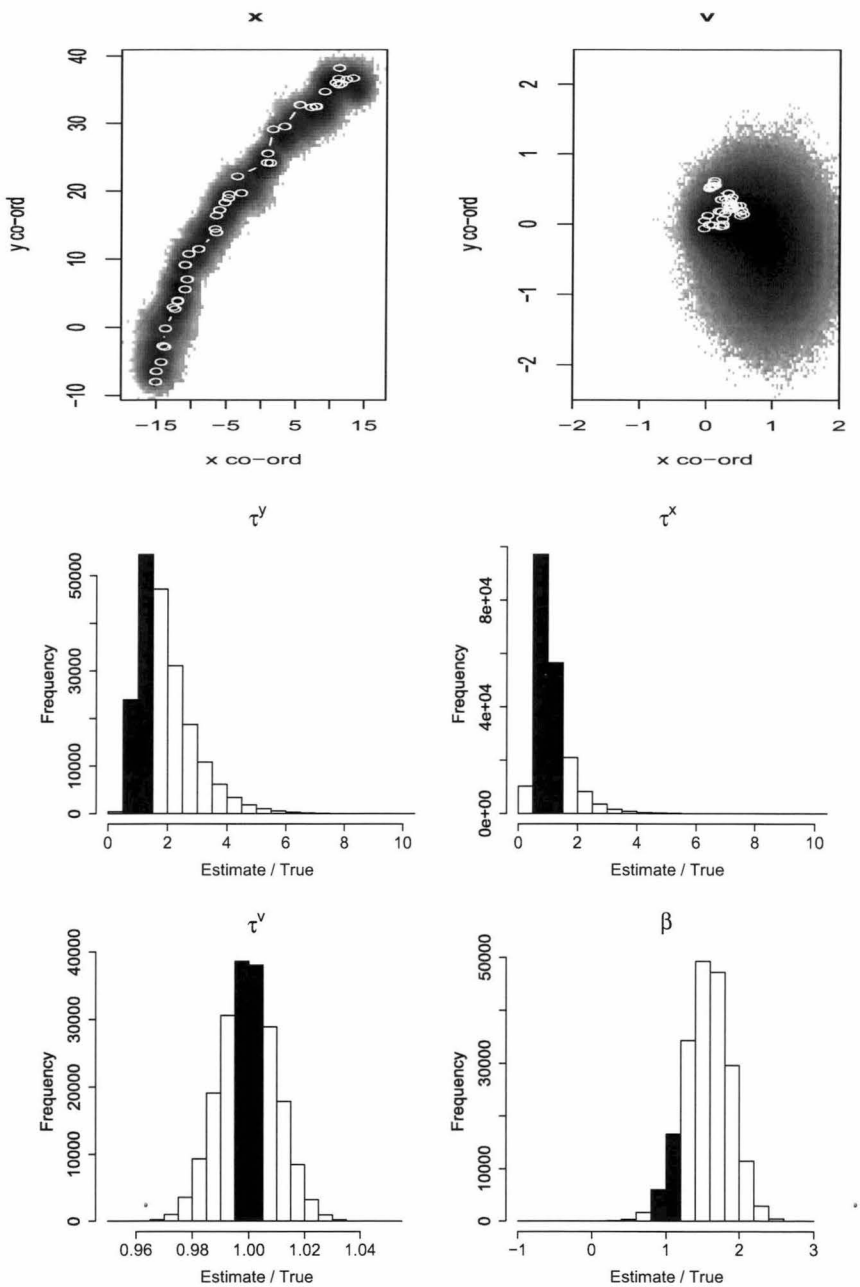


Figure 3.9: Results from 200 000 MCMC draws from the posterior of the estimation model that uses northerly advection (scenario 1), applied to the track simulated using a combination of northerly advection and a gyre (scenario 4). States are shown as Cartesian coordinates (the simulated values as open circles joined by white lines and the MCMC draws as a density plot). Parameters are shown as a fraction of their true values so that a correct value is given as 1; the bars that span this value are shaded in black.

to a wide variety of forms of external as well as internal factors. This framework includes the estimation of latent locations so that the time resolution of the model may be increased when gaps exist in the track data. An illustration, drawn from the family of possible forms of this model, is used to show that, in simulation, the model correctly identifies parameter values and states, and this was found generally.

In this framework, combinations of advection forces are used to represent hypotheses regarding movement behaviour. These are estimated using co-models that may use environmental data, such as SST fields, or may simply use theory, such as a tendency to travel northwards. Alternative co-models, representing alternative behavioural hypotheses, may use the same environmental data but different behaviours. For example, two co-models may both use SST data but one may reflect a tendency to move towards warmer water whereas the other reflects avoidance thereof. Different hypotheses of movement behaviour might lead to similar advection fields so that the method presented here might not always be able to choose among competing hypotheses. Similarly, a single idea such as that 'white sharks navigate using the earth's magnetic field' does not lead to a single advection field for a shark that is moving towards a known location using the geomagnetic field to guide it there. Would it be able to move straight there by the shortest route or would it curve in towards its goal due to non-orthogonality of the components of the geomagnetic field? Would it need to swim for a long distance in an arbitrary direction in order to sample the geomagnetic field before it could orientate? Each suggested answer to these questions would result in a different advection field so that from the point of view of someone using this modelling framework a 'hypothesis' of movement behaviour is a complete set of suggestions that leads to a single advection field.

If advection fields used in combination in the same model are correlated with one another in the region through which the animal moves, then confounding will occur. Care should be taken to avoid this, unless the animal also moves through regions where these fields operate in independent directions.

Inference regarding which advection forces are influencing movement behaviour can be made by including several advection forces in the model and then examining the estimated coefficients β which give an indication of the relative effect of each force. A zero estimate for one of the coefficients would indicate that the corresponding advection force is not effecting movement. However, this will only be possible when the advection fields used are not correlated in the areas through which the animal moves. As shown in Chapter 5, further inference can be made by constructing models that use different advection forces or combinations of advection forces and using

the Bayes factor to assess which best explain the observed track.

The model is unable to estimate the precision parameters with accuracy, seeming to rely on the influence of the prior. However, this does not lead to serious bias in the estimates of other model parameters. However, this might affect the results of model comparison, a problem that is explored in Chapter 5.

The precision parameter for velocity τ^v is largely uncorrelated with those for observation τ^y and location τ^x. Its value influences the precision of the marginal for the advection coefficient β but has little influence on its location. In this investigation, the variance $= \text{diag}(1, 1)$ assumed for the prior for τ^v was the same as that used for the other precision terms τ^y and τ^x. However, because a much smaller mean value was assumed, there was less room for variation in the resulting marginal distribution τ^v. This parameter influences the inertia, or smoothness, of the path, governing the rate at which the direction of inertia can change. The inertial property of the path is considered to be part of the model selection choice made by the investigator. Thus τ^v is considered to be a model setting, rather than a parameter whose estimated value is of interest, so that a more precise choice for the prior for τ^v is acceptable. This, along with many other choices of model framework may, of necessity, be relatively arbitrary. Guidance, in this case, would come from the investigator's assessment of the directedness of the animal. Is the animal meandering aimlessly (imprecise velocity, a small value for τ^v) or is it moving directly towards an apparent goal (precise velocity, a large value for τ^v)?

A shift in the centre of the posterior marginal for v, away from the origin, might indicate that the advection fields used in the estimation model are poor representations of the movement behaviour of the tracked animal. This indicates a directional persistence that has not been accounted for by the chosen advection fields, so that an additional motivation for movement exists that has not been included in the estimation. Investigation of further instances of incorrect choice of advection fields is desirable to test the robustness of this indicator of incorrect choice. Such examinations were beyond the scope of this chapter. The primary aim of this thesis is to derive an effective method of model choice, rather than an accurate estimator of parameter values; therefore Chapter 5 presents investigations into the robustness of methods of model selection to incorrect model specification, including incorrect choice of advection field.

This chapter's application of a discrete time model to a process that is continuous in time, is an approximation. Alternative choices are available for the way in

which velocity v and the precision terms τ^x and τ^v were scaled with time. The Brownian motion option used here seemed to be the most natural. In principle, an exact solution would require methods such as stochastic differential equations (Jones & Ackerson, 1990), but it is not clear that the accuracy of the data likely to be available will warrant this level of time resolution. However, the ability to include latent locations does allow some control over the time resolution of the model. Typical tracking data include occasional long periods without position estimates and if these occur when the animal is passing through a region where the advection field undergoes significant change of direction then latent locations become necessary in order to describe the likely route taken. Nevertheless, when developing behavioural co-models, care should be taken to avoid algorithms that give rise to advection fields whose direction changes greatly over a short distance, as this will result in a movement model that is sensitive to small errors in location. Such an example is given in Chapter 8 where the bathymetric observations used to generate an advection field were smoothed prior to use.

The use of a Student-t distribution to describe observation errors would provide robust inference in the presence of outlying observations (Gelman *et al.* , 2004). Jonsen *et al.* (2005) and Flemming *et al.* (2006) have demonstrated the use of t-distributed observation errors in state space models of movement behaviour. The purpose of this Chapter is to emphasize other, novel, aspects of this modelling framework, therefore a more tractable, Gaussian, error structure was chosen for this illustration.

Jonsen *et al.* (2003) modelled the tendency for an animal to move in a preferred direction by allowing non-zero covariance terms for their process error. This necessitated the use of a Wishart distribution, less tractable than the gamma distributions used here. Directional preference is achieved more easily by means of advection.

Many animals show inertia, or directional persistence, a tendency to continue moving in the same direction (Turchin, 1998), which it has been suggested, indicates navigational uncertainty (Guilford *et al.* , 2004; Roberts *et al.* , 2004). This is represented here by adding one level (v) of stochastic trend to the structural time series that is the state equation, thus adding greater autocorrelation to the track. More levels could be added (Chatfield, 2004; Harvey, 1990) but these overcome the effect of advection to some degree, more strongly the more levels added. Similarly, greater precision in the errors in velocity (larger τ^v) results in greater autocorrelation in the path. The choice of a single additional layer and relatively precise τ^v gave paths that appeared consistent with the satellite tracks of white sharks (see Chapter 8).

Stochastic processes were applied to both the path x and the velocities v, but one of these error terms could have been omitted. Using both, simplified the mathematics and emphasized the state space nature of the model. However, it reduced the identifiability of the model because an additional precision parameter, and several additional states, had to be estimated. It might be preferable, in future, to remove one of these sources of error.

Priors $P(x_1 \mid \theta)$ and $P(v_1 \mid \theta)$ were imposed on the initial conditions for both path x and velocity v. However, the initial location of the tagged animal is likely to be known with sufficient accuracy to warrant fixing the value of x_1. It is less likely that v_1, the velocity (additional to that given by advection) and inertial direction prior to capture, will be known, and even if it were, the trauma of capture and release might cause a change. However, given the model's poor ability to estimate v, the value assumed for v_1, or the degree of uncertainty assumed in its prior, are unlikely to affect the conclusions drawn from the modelling study.

The 'conditional' for x (equation 3.24) is a good proposal distribution for x provided X does not show great variation over small distances (so that it is not locally smooth at the scale of the errors in the observations ϵ and the true locations ξ). It was found to be an adequate proposal distribution for the simulations presented in this thesis, but if less smoothly varying advection fields are used then the MCMC chain for x may not mix well. If this occurs, it will appear in trace plots of MCMC draws for x as periods during which the values of x do not change, because of repeated rejection of the proposed values. In such a case, it might be wise to consider alternative proposal distributions for x. These could be bivariate normals around the observed locations y, or around the x values from the previous MCMC step. Alternatively, the 'conditional' could be used as the proposal distribution but with the variance decreased by some constant. This retains the covariance structure of the 'conditional' but decreases the overall variance, thus reducing the proposed x value's distance from the mean of the proposal distribution and therefore increasing the chance that it lies within a region in which the advection field is locally constant. Alternatively, an appropriate nonlinear version of the Kalman filter could be used to estimate the state variables (Wang, 2007; Carter & Kohn, 1994). Note that when advection is independent of x (such as when the advection field has the same value at all locations) the Metropolis-Hastings within Gibbs scheme will reduce to Gibbs sampling, as illustrated by equation 3.28.

Ideally, in each of the scenarios outlined in this (and subsequent) chapters sev-

eral tracks would have been simulated and the results of model application to all tracks compiled. However, the computational burden of this work precludes such care. The presentation of several, different, scenarios across several chapters offers, instead, a degree of replication. During development several tracks were examined and the results presented here are typical.

Chapter 4

Changing behaviour through time

4.1 Abstract

Animal movement behaviour is subject to change over time, sometimes over the course of a single tracking period. In this chapter, the modelling framework presented in the previous chapter is extended to allow behaviour to change with time. This is achieved by allowing the coefficients β for the advection fields to be dynamic. The model allows for correlation in the changes for all coefficients so that, for example, one behaviour can become dominant while another wanes.

The change in the advection coefficients is allowed to follow a random walk. The estimated coefficient values are best at the middle of the tracking period, as the earliest and latest values are least constrained. There is considerable scope for confounding among advection coefficients. This could be curbed to some degree by using a truncated prior to prevent the estimation of negative values for the coefficients.

4.2 Introduction

Animal movement behaviours are subject to change over time. Changes may be cyclic, such as breeding behaviours and responses to seasonal changes in the abundance and distribution of food sources. Alternatively changes may be in response to the environment, such as prey-switching behaviour when one food source becomes rare relative to others (Ostfeld, 1982). Others may be permanent changes such as those that occur as an animal matures e.g. breeding areas will be visited only after the onset of maturity (Manning & Dawkins, 1992). It is unlikely that long-term

changes such as those associated with maturity would be of great relevance over the time scale of track data, neither is a whole seasonal cycle likely to be evident, however the onset of such changes, over a period of weeks or months, is likely to be evident at the scale of a track. Therefore, it is desirable to allow modelled behaviour to change over time. It is also desirable to correlate the change in behaviours because it is conceivable that as one behaviour becomes more important another may fade (Manning & Dawkins, 1992). For example, during the onset of a mating season, the desire to find good foraging grounds may weaken while the desire to move to a spawning ground simultaneously increases.

The model presented in Chapter 3 represents movement behaviour by means of advection fields that direct the movement of a tracked individual. A change in behaviour over time could be represented in one of three ways. First, the advection field could change with time in a pre-specified manner. Here, the co-model is run first, before the movement model, so that the specified change in advection, forms part of the hypothesis regarding movement behaviour. The parameters of the co-model that give rise to the advection field are fixed, not estimable model parameters. This is not a flexible way of modelling changing behaviour.

Second, the advection co-model could be integrated into the movement model so that they run simultaneously. Here, the parameters of the co-model are included in the estimated set of model parameters. This is likely to greatly increase the computational burden and might lead to problems of estimability due to confounding of parameters, but it would be a very flexible means of estimating movement behaviour. Considerable further work would be required to investigate and develop the properties of such a model.

Third, the model of the previous chapter can be extended to allow the coefficients of the advection fields to change with time. Each field has an associated coefficient that governs how strongly that field influences movement. A value of zero indicates that the field does not affect behaviour at all, a larger value indicates a stronger influence, and negative values cause the animal to move against the advection flow. In Chapter 3, the values of these coefficients are estimated by the model, and are then held fixed for all time steps so that the relative adherence of the moving animal to each advection field does not change over time. This is a more flexible approach than that of incorporating the behavioural change into a pre-specified advection co-model, and is less computationally burdensome and possibly more stable than incorporating the co-model into the movement model. Such state space models with dynamic coefficients (linear or nonlinear dynamic models) were the subject of West

& Harrison (1997)'s work.

This chapter presents a version of the model presented in Chapter 3 that has dynamic advection coefficients so that behaviour can change over time. The change in the coefficients can be correlated so that the importance of one behaviour may increase while another, correspondingly, fades.

4.3 Methods

4.3.1 The model

A full explanation of the movement modelling framework is given in Chapter 3. The current chapter concentrates on presenting the changes that result from allowing that model's advection coefficient parameter β to change over time. Symbols used in this chapter are listed in Appendix 10 along with a brief description, and their dimensionality.

To allow the response to advection to change over time, the set of advection coefficients β (Chapter 3) are replaced by B_k, the vector of coefficients specific to time step k , $k = 1 \ldots N-1$. In turn, these are assumed to undergo a random walk (for reasons that are discussed later, a functional form is not imposed on this change)

$$B_{k+1} = B_k + \eta_k, \qquad k = 1 \ldots N-2. \tag{4.1}$$

The errors η_k are assumed to be independent between time steps. However, the changes occurring in the m advection coefficients may be correlated with one another - that is, the errors η_k have a multivariate normal distribution

$$\eta_k \sim \text{MVN}_m\left(0, \Delta_k \left(\tau^B\right)^{-1}\right) \tag{4.2}$$

with possibly non-diagonal variance matrix $(\tau^B)^{-1}$. Non-zero covariance terms would result in correlation between the changes in the coefficients. Because the change in behaviour occurs as a function of time and not time step, the $m \times m$ variance term $(\tau^B)^{-1}$ scales with the duration of the time step ($\Delta_k = t_{k+1} - t_k$) (equivalent to Brownian motion).

The state equation for the path (previously equation 3.3) is now

$$x_{k+1} = x_k + \Delta_k \left(v_k + B_k X_k^T\right) + \xi_k, \qquad k = 1 \ldots N-1. \tag{4.3}$$

The random walk for velocity, the term that introduces inertia to the path, is unaltered

$$v_{k+1} = v_k + \zeta_k, \qquad k = 1 \ldots N-2. \tag{4.4}$$

Recall that $x_k = (x_{k1}, x_{k2})$ and $v_k = (v_{k1}, v_{k2})$, representing the two Cartesian co-ordinates of location, and that the advection matrix X_k has dimension $2 \times m$ representing the m advection fields thought to influence movement behaviour, in each of the two Cartesian directions (Section 3.3.2).

The observation equation 3.1 for y_k is unaltered

$$y_k = x_k + \epsilon_k \quad k \in O, \tag{4.5}$$

and so are the error distributions for the observations

$$\epsilon_k \sim \mathrm{N}\left(0, \mathrm{diag}\,(\tau_1^y, \tau_2^y)^{-1}\right) \quad k \in O, \tag{4.6}$$

for the path

$$\xi_k \sim \mathrm{N}\left(0, \Delta_k \ \mathrm{diag}\,(\tau_1^x, \tau_2^x)^{-1}\right) \quad k = 1 \ldots N-1, \tag{4.7}$$

and for inertia

$$\zeta_k \sim \mathrm{N}\left(0, \Delta_k \ \mathrm{diag}\left(\tau_1^v, \tau_2^v\right)^{-1}\right) \qquad k = 1 \ldots N-2. \tag{4.8}$$

4.3.2 Priors

Gamma priors were specified for the scalar precision terms for the errors in location, inertia and observation (see Section 3.3.4) corresponding with each of the Cartesian directions. However, the precision for the errors in the advection coefficients τ^B is not scalar because correlation has been allowed for. A conjugate Wishart prior (Diaconis & Ylvisaker, 1979) is adopted for the precision τ^B for the errors in the advection coefficients B_k. The Wishart can be thought of as a multivariate equivalent of the gamma distribution (Gelman *et al.* , 2004; Evans *et al.* , 2000)

$$\tau^B \sim \mathrm{Wishart}_m(n_0, S_0). \tag{4.9}$$

Because the advection coefficients B, which were model parameters in Chapter 3, now follow a random walk with estimable error parameters, B now joins x and v as an unobserved state vector. Consequently, a prior must be specified for the advection coefficients in the first time step B_1, like those for x_1 and v_1 (Section 3.3.4). A multivariate normal prior is used, with mean B_0 and 2×2 precision matrix τ^{B0}, which may have non-zero covariate terms

$$B_1 \sim \mathrm{MVN}_2\left(B_0, \ \left(\tau^{B0}\right)^{-1}\right). \tag{4.10}$$

4.3.3 The posterior

The coefficients B must now be included in the posterior (previously equation 3.7) in terms of their dependence on the parameter vector θ, which now represents only the precision parameters τ^y, τ^x, τ^v, and τ^B

$$p(x, v, B, \theta \mid y) \propto p(y \mid x, v, B, \theta)\ p(x \mid v, B, \theta)\ p(v \mid B, \theta)\ p(B \mid \theta)\ p(\theta). \tag{4.11}$$

4.3.4 Conditional distributions

Once again (as in Section 3.3.5) having used conjugate priors for the parameters, full conditional distributions can be calculated for all parameters and states with the exception of x when X depends on x. These conditional distributions are shown below. Their method of calculation is detailed in Appendix 11. Detailed calculations for a gamma conditional are shown in Appendix 12, for a multivariate normal conditional in Appendix 13, and for the Wishart conditional distribution for τ^B in Appendix 14.

First, the error terms are expressed in vector form.

The path location errors ξ_k at each time step $k = 1 \ldots N-1$ are a function of the product of m advections for each Cartesian co-ordinate $[\mathcal{X}_{k1} \mid \mathcal{X}_{k2}]$ with m corresponding advection coefficients B_k. In order to express this in vector notation, the partitioned matrix $\left[\tilde{\mathcal{X}}_1 \mid \tilde{\mathcal{X}}_2\right]$ is defined such that $\tilde{\mathcal{X}}_i$ is an $(N-1) \times m(N-1)$ matrix consisting of m blocks of $(N-1) \times (N-1)$ diagonal matrices. The diagonal elements of the mth block are the $N-1$ advection terms of the mth advection field for coordinate i.

The product $\tilde{\mathcal{X}}_i$ vec(B) is an $N-1$ column vector of weighted sums (equivalent to $[\mathcal{X}_1 \mid \mathcal{X}_2]\,\beta$ from equation 3.17, for each time step). Here, 'vec' is the vectorization operator (Harville, 1997), which stacks the columns of a matrix. The path location errors ξ are given by

$$\xi = D_x\, x - \Lambda_x \left(v + \left[\tilde{\mathcal{X}}_1 \text{ vec}(B) \mid \tilde{\mathcal{X}}_2 \text{ vec}(B)\right]\right). \tag{4.12}$$

As before, D_x is a matrix that gives the $N-1$ first differences of x, and Λ_x is a diagonal matrix of step durations (Section 3.3.5).

The errors in the random walk for the advection coefficients B is given by

$$\eta = D_v\, B \tag{4.13}$$

where D_v is a matrix that gives the $N-2$ first differences for each column of any matrix that has $N-1$ rows.

The observation errors are unchanged; equation 3.15 is repeated

$$\epsilon = y - P\,x, \tag{4.14}$$

as are the errors in velocity; equation 3.18 is repeated

$$\zeta = D_v\,v. \tag{4.15}$$

When stating the full conditional distributions for the parameters, $\theta(-\tau_i^y)$ is used to indicate all parameters except τ_i^y. Once again V represents the totality of environmental and behavioural information to which the individual may respond and is not, therefore, dependent on location x (Section 3.3.4).

The conditional for the precision matrix for errors in the advection coefficients τ^B is Wishart

$$\tau^B \mid \theta(-\tau^B), x, v, y, B, V \sim \text{Wishart}\left(n_0 + N - 2,\ S_0^{-1} + \eta^T \Lambda_v^{-1} \eta\right). \tag{4.16}$$

The conditional for the $m(N-1) \times 1$ vector of advection coefficients $\text{vec}(B)$ is a multivariate normal

$$B \mid \theta, x, v, y, V \sim \text{MVN}_{\text{m(N-1)}}\left(\mu^B,\ \Sigma^B\right), \tag{4.17}$$

with mean

$$\mu^B = \Sigma^B \left\{ \sum_{i=1}^{2} \left[\tau_i^x\ \tilde{\mathcal{X}}_i^T\ (D_x\,x_i - \Lambda_x v_i) \right] + Q_B\ \tau^{B0}\ B_0^T \right\}$$

and variance

$$\Sigma^B = \left[Q_B \, \tau^{B0} \, Q_B^T + D_B^T \left(\Lambda_v^{-1} \otimes \tau^B \right) D_B + \sum_{i=1}^{2} \tau_i^x \, \tilde{\mathcal{X}}_i^T \, \Lambda_x \, \tilde{\mathcal{X}}_i \right]^{-1} .$$

Here Q_B is an $m(N-1) \times m$ matrix whose upper left $m \times m$ elements are the identity matrix of dimension m, and whose remaining elements are zero. This matrix operates on the parameters of the prior for the initial state B_1, increasing their dimensions to those of B by the addition of zeros.

The matrix D_B returns the $m(N-2)$ first differences of the advection coefficients B such that $D_B \ \mathrm{vec}(B) = \mathrm{vec}(D_v \, B)$.

The dimensions of all these quantities are listed in Appendix 10.

The states v_i, $i = 1, 2$ have a multivariate normal conditional distribution

$$v_i \mid \theta, x, v_{-i}, y, B, V \sim \mathrm{MVN}_{\text{N-1}} \left(\mu_i^v, \Sigma_i^v \right) \tag{4.18}$$

with mean

$$\mu_i^v = \Sigma_i^v \left[\tau_i^x \left(D_x \, x_i - \Lambda_x \, \tilde{\mathcal{X}}_i \ \mathrm{vec}(B) \right) + \tau_i^{v0} \, v_i^0 \, \mathcal{I}_{v0} \right]$$

and variance

$$\Sigma_i^v = \left[\tau_i^x \, \Lambda_x + \tau_i^v \, D_v^T \, \Lambda_v^{-1} \, D_v + \tau_i^{v0} \, \mathcal{I}_{v0} \, \mathcal{I}_{v0}^T \right]^{-1}$$

where $\mathcal{I}_{v0}$ is a column vector of length $N-1$ of the form $(1, 0, 0, ..., 0)^T$, which expands the parameters of the prior for the initial condition v_1 by adding zeros.

If X were independent of x then the conditionals for each x_i, $i = 1, 2$ would be multivariate normal distributions of order N (the total number of locations including latent locations)

$$x_i \mid \theta, x_{-i}, v, y, B, X \sim \text{MVN}_\text{N}\left(\mu_i^x,\ \Sigma_i^x\right) \tag{4.19}$$

with mean

$$\mu_i^x = \Sigma_i^x \left[\tau_i^y\ P^T\ y_i + \tau_i^x\ D_x^T\ (\tilde{X}_i\ \text{vec}(B) +\ v_i) + \tau_i^{x0}\ x_i^0\ \mathcal{I}_{x0}\right]$$

and covariance

$$\Sigma_i^x = \left[\tau_i^y\ P^T\ P + \tau_i^x\ D_x^T\ \Lambda_x^{-1}\ D_x + \tau - i^{x0}\ \mathcal{I}_{x0}\ \mathcal{I}_{x0}^T\right]^{-1},$$

where $\mathcal{I}_{x0}$ is a column vector of length N of the form $(1, 0, 0, ..., 0)^T$, which is used to expand the parameters of the prior distribution for the first location x_1 to those of x.

The conditional for the precision for the observation errors has not changed

$$\tau_i^y \mid \theta(-\tau_i^y), x, v, y, B, V \sim \text{Gamma}\left(\frac{N}{2} + a_i^y,\ \frac{1}{2}\ \epsilon_i^T\ \epsilon_i + b_i^y\right) \tag{4.20}$$

and neither has that for the precision for the path location errors, although these errors ξ_i were calculated (by equation 4.12) using the new matrix product for the advections

$$\tau_i^x \mid \theta(-\tau_i^x), x, v, y, B, V \sim \text{Gamma}\left(\frac{N-1}{2} + a_i^x,\ \frac{1}{2}\ \xi_i^T\ \Lambda_x^{-1}\xi_i + b_i^x\right). \tag{4.21}$$

The conditional for the inertial state is unaltered

$$\tau_i^v \mid \theta(-\tau_i^v), x, v, y, B, V \sim \text{Gamma}\left(\frac{N-2}{2} + a_i^v,\ \frac{1}{2}\ \zeta_i^T\ \Lambda_v^{-1}\ \zeta_i + b_i^y\right). \tag{4.22}$$

4.3.5 Simulation and estimation

A set of advection scenarios was chosen in order to simulation test this estimation model. Because correlations are allowed between advection fields, a minimum of two advection fields is needed to fully test the model. The simplest advection fields are constant, that is, the advection at all points is the same. A more complicated advection field, a circular gyre, is also considered in order to test the ability of the model to distinguish a constant advection field from one that changes direction with location. Two alternative pairs of simple advection fields are considered: first, a constant northerly flow coupled with a northeasterly flow; second, a constant southerly flow coupled with an anticlockwise gyre (Figure 4.1).

Initially, only one of the advection fields is important, but over time, adherence to that one diminishes as the other increases. Two likely functional forms were chosen for the change in the advection coefficients B: one in which the changes occur linearly, and another in which they follow an S-shaped curve so that no change occurs for a period followed by a period of rapid change and then another without change (Figure 4.1).

These four scenarios are detailed in Table 4.1 and the parameter values used to simulate tracks are shown in Table 3.1. The changes in the advection coefficients B are assumed to occur smoothly with time, not with time step. Therefore, because time steps differ in duration, the plots of B against time step (Figure 4.1) are not smooth.

Table 4.1: Description of four advection scenarios that use different functional forms for the changes in the advection coefficients B with time, and different advection fields.

Name	Change in B	Advection Fields
Scenario 1	Linear	North; Northeast
Scenario 2	Linear	South, Gyre
Scenario 3	S-shaped	North; Northeast
Scenario 4	S-shaped	South, Gyre

Estimation is by MCMC, as described in Chapter 3 and the priors are given in Table 4.3. Because this investigation focuses on the estimation of changing advection coefficients B, the simulation used more precise (larger) values for the precision parameters for the observations τ^y, and for the path τ^x than those used in Chapter 3. The same, precise, value was used for the precision for velocity τ^v and which again

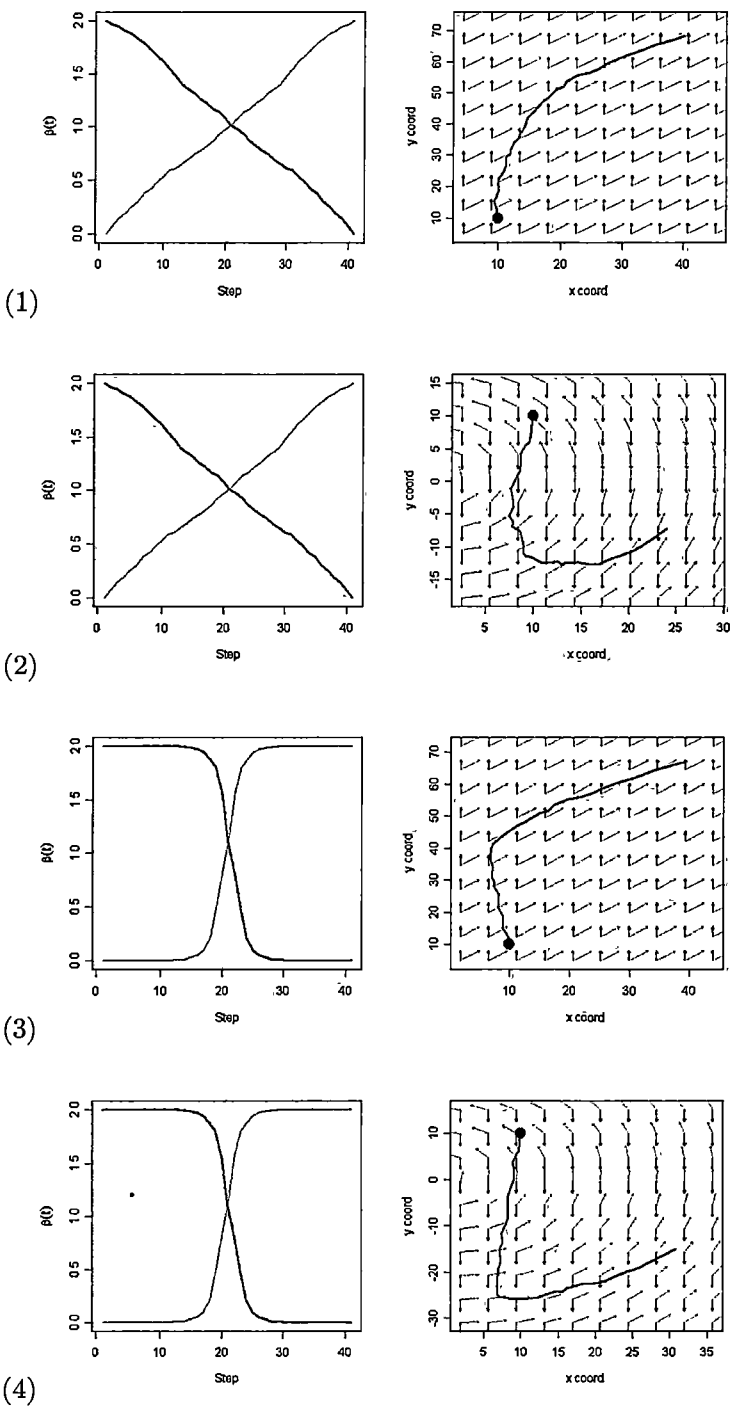


Figure 4.1: B values and corresponding advection fields used in the four simulation scenarios. The Bs drawn as black lines correspond with the advection forces shown as black arrows, and the greys similarly. The simulated track is overlaid; its starting point shown as a black dot

Table 4.2: Parameter values and dimensions used to simulate tracks.

Symbol	Description	Value
N	Number of path locations x	41
n	Number of observed locations y	21
m	Number of advection fields	2
x_1	Initial location	$(-10,\ -10)$
v_1	Initial velocity	$(0, 0)$
Λ	Step durations (drawn randomly)	$\sim$ N(0.75, 0.3)
X_k	Advection forces	see Table 4.1
B	Coefficient of advection	see Table 4.1
O	Index for observed locations	$1, 3, \ldots 41$
τ^y	Precision for observed locations y	$(0.1^{-1},\ 0.1^{-1})$
τ^x	Precision for locations x	$(0.1^{-1},\ 0.1^{-1})$
τ^v	Precision for velocity v	$(0.01^{-1},\ 0.01^{-1})$
B_1	Initial advection coefficients	$(1,\ 0)$

reflected the idea that this is regarded as a model setting rather than a parameter to estimate. Priors with relatively diffuse variances ($= 1$) were used for all precision terms.

Scenarios 1 and 4 use constant advection fields so that if these are used in the estimation model, then the MCMC routine will be at its most efficient, because it will revert to Gibbs sampling. Scenarios 2 and 4 use one non-constant advection field so that it is necessary to use a Metropolis-Hastings step to estimate the path x.

For each of the four scenarios 100 000 MCMC draws were made. Convergence was indicated (although not guaranteed) using the CODA diagnostic software package.

4.4 Results

The posterior marginal for velocity shows roughly half the spread seen in Chapter 3, (Figures 4.2 to 4.5) because greater precision was used when simulating the path (larger τ^x). Once again, the locations estimates are relatively good.

A persistent, but small, downwards bias is seen in the posterior marginal for the precision for the errors in the path τ^x (Figures 4.2 to 4.5). An even smaller, but again persistent, downward bias is seen in the corresponding precision for velocity τ^v.

Table 4.3: Priors used in estimation. Note that the shape a and rate b parameters of the gamma distribution can be calculated from the distribution's mean μ and variance σ^2 using the formulae $a = \mu^2/\sigma^2$ and $b = \mu/\sigma^2$. The gamma prior distributions are expressed both in terms of shape a and rate b, and also mean and variance.

Symbol	Description	Value
x^0	Mean initial location	$(-10, -10)$
v^0	Mean initial velocity	$(0, 0)$
τ^{x0}	Precision for initial location	$(0.001^{-1}, 0.001^{-1})$
τ^{v0}	Precision for initial velocity	$(0.001^{-1}, 0.001^{-1})$
B_0	Mean initial advection coefficients	$(0, 0)$
τ^{B0}	Precision for prior for B_1	$\text{diag}(1, 1)$
n_0	Degrees of freedom for Wishart prior for τ^B	2
S_0	Scale matrix for Wishart prior for τ^B	$\begin{pmatrix} 2.5 & 2 \\ 2 & 2.5 \end{pmatrix}$
a^y	Shape for gamma prior for τ^y	$(0.01^{-1}, 0.01^{-1})$
b^y	Rate for gamma prior for τ^y	$(0.1^{-1}, 0.1^{-1})$
	Mean for gamma prior for τ^y	$(0.1^{-1}, 0.1^{-1})$
	Variance for gamma prior for τ^y	$(1, 1)$
a^x	Shape for gamma prior for τ^x	$(0.01^{-1}, 0.01^{-1})$
b^x	Rate for gamma prior for τ^x	$(0.1^{-1}, 0.1^{-1})$
	Mean for gamma prior for τ^x	$(0.1^{-1}, 0.1^{-1})$
	Variance for gamma prior for τ^x	$(1, 1)$
a^v	Shape for gamma prior for τ^v	$(1e4, 1e4)$
b^v	Rate for gamma prior for τ^v	$(0.01^{-1}, 0.01^{-1})$
	Mean for gamma prior for τ^v	$(0.01^{-1}, 0.01^{-1})$
	Variance for gamma prior for τ^v	$(1, 1)$

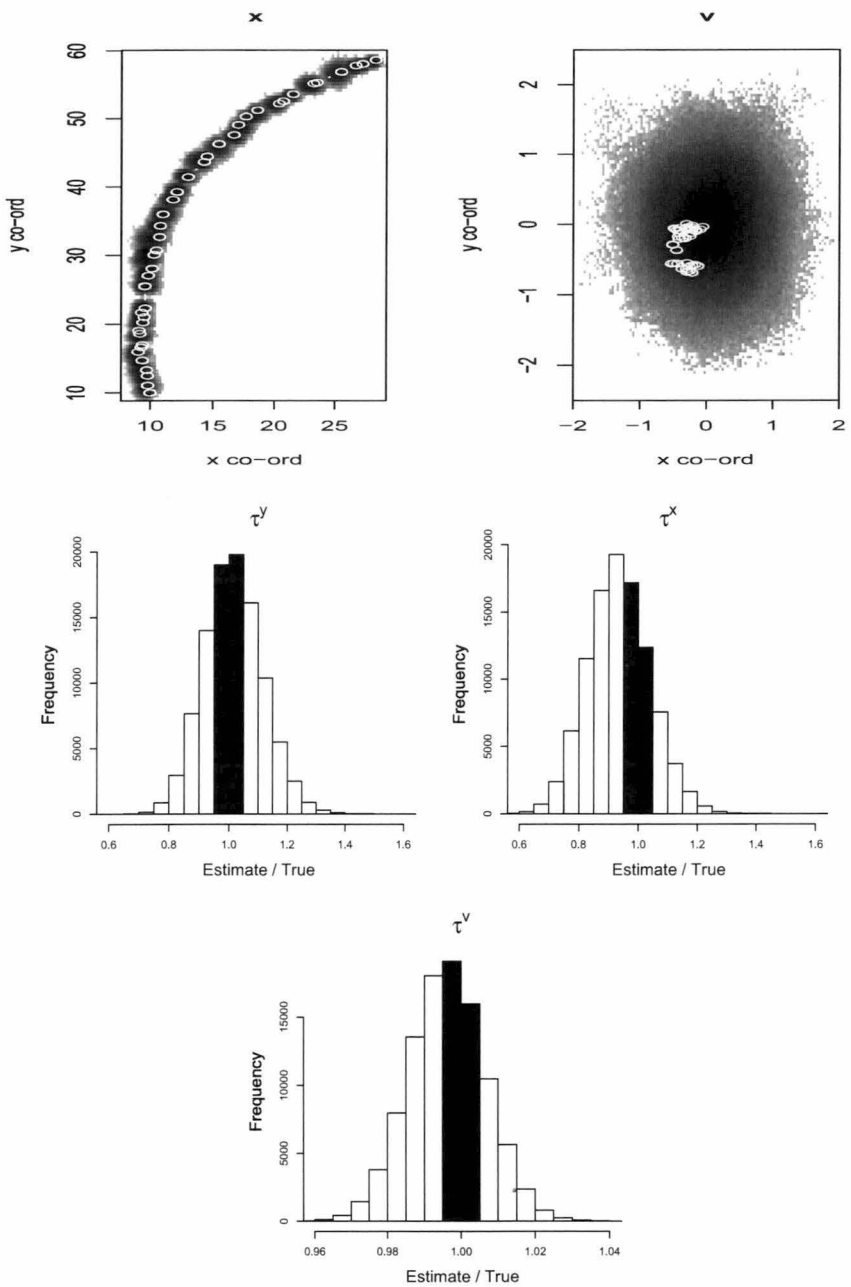


Figure 4.2: Estimation results for scenario 1, which uses linearly declining adherence to a northerly advection field combined with linearly increasing adherence to a northeasterly field. The x and v plots show the true values as white, open circles and the 100 000 MCMC draws as a density plot. The precision parameters τ^y, τ^x and τ^v are shown as a fraction of their true values, so that a correct value is given as 1; the bars that span this value are shaded in black.

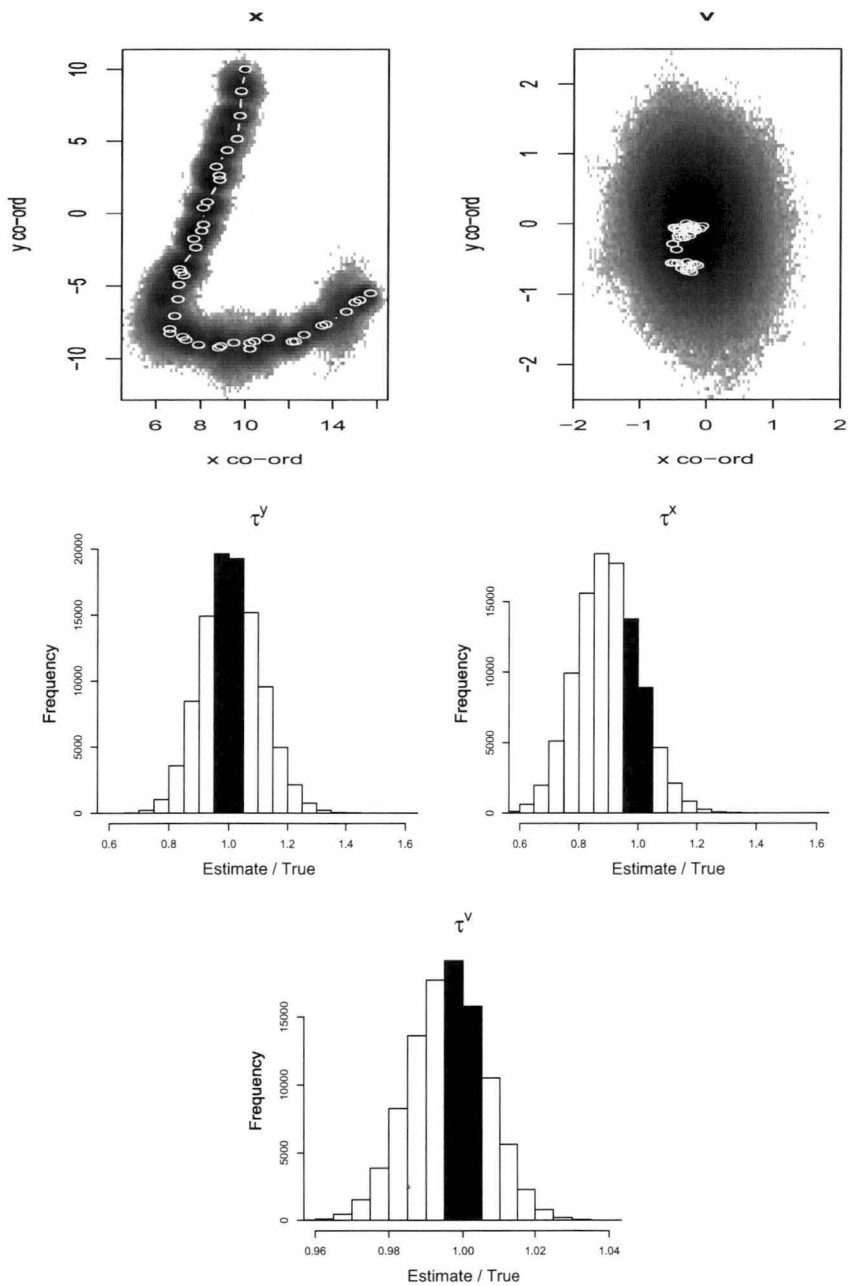


Figure 4.3: Estimation results for scenario 2, which uses linearly declining adherence to a southerly advection field combined with linearly increasing adherence to a gyre. The x and v plots show the true values as white, open circles and the 100 000 MCMC draws as a density plot. The precision parameters τ^y, τ^x and τ^v are shown as a fraction of their true values, so that a correct value is given as 1; the bars that span this value are shaded in black.

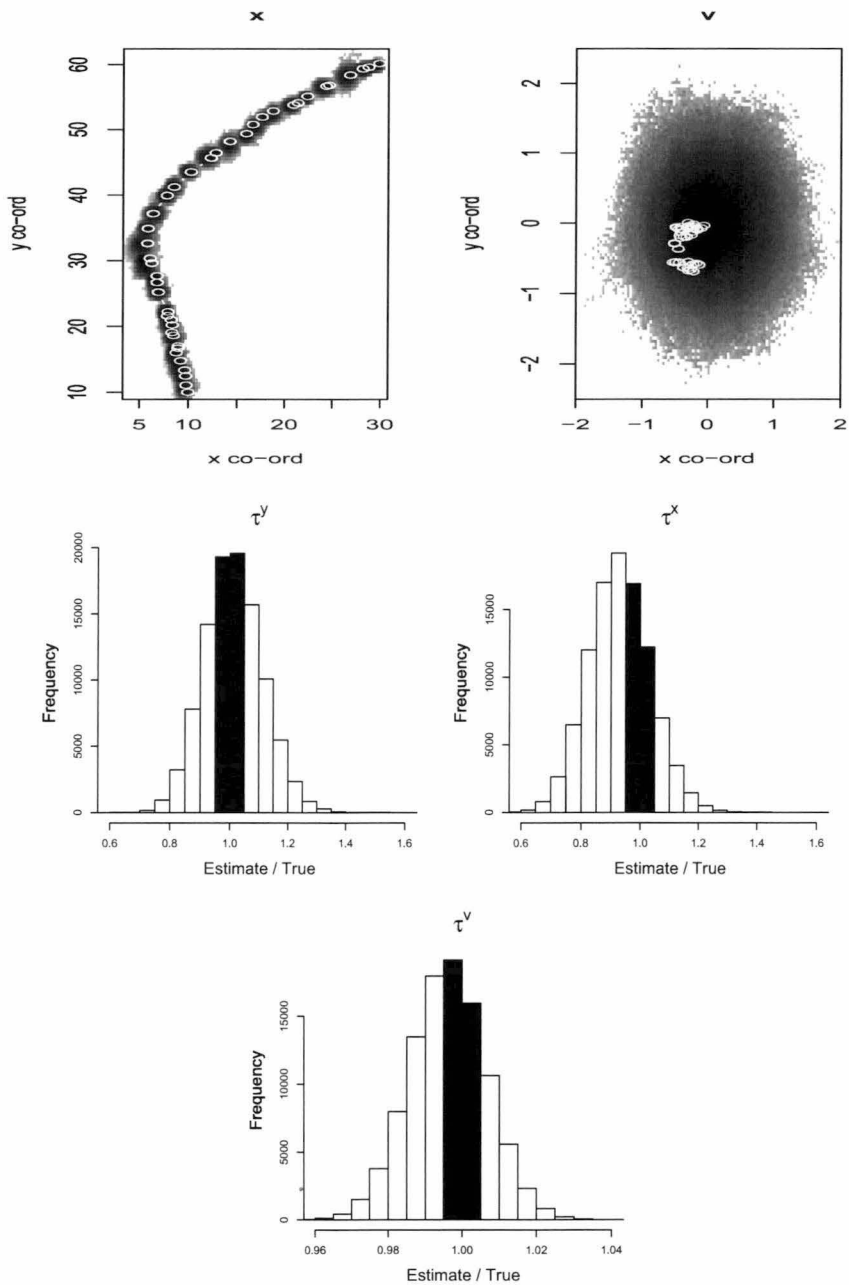


Figure 4.4: Estimation results for scenario 3, which uses an S-shaped declining adherence to a northerly advection field combined with an S-shaped increasing adherence to a northeasterly field. The x and v plots show the true values as white, open circles and the 100 000 MCMC draws as a density plot. The precision parameters τ^y, τ^x and τ^v are shown as a fraction of their true values, so that a correct value is given as 1; the bars that span this value are shaded in black.

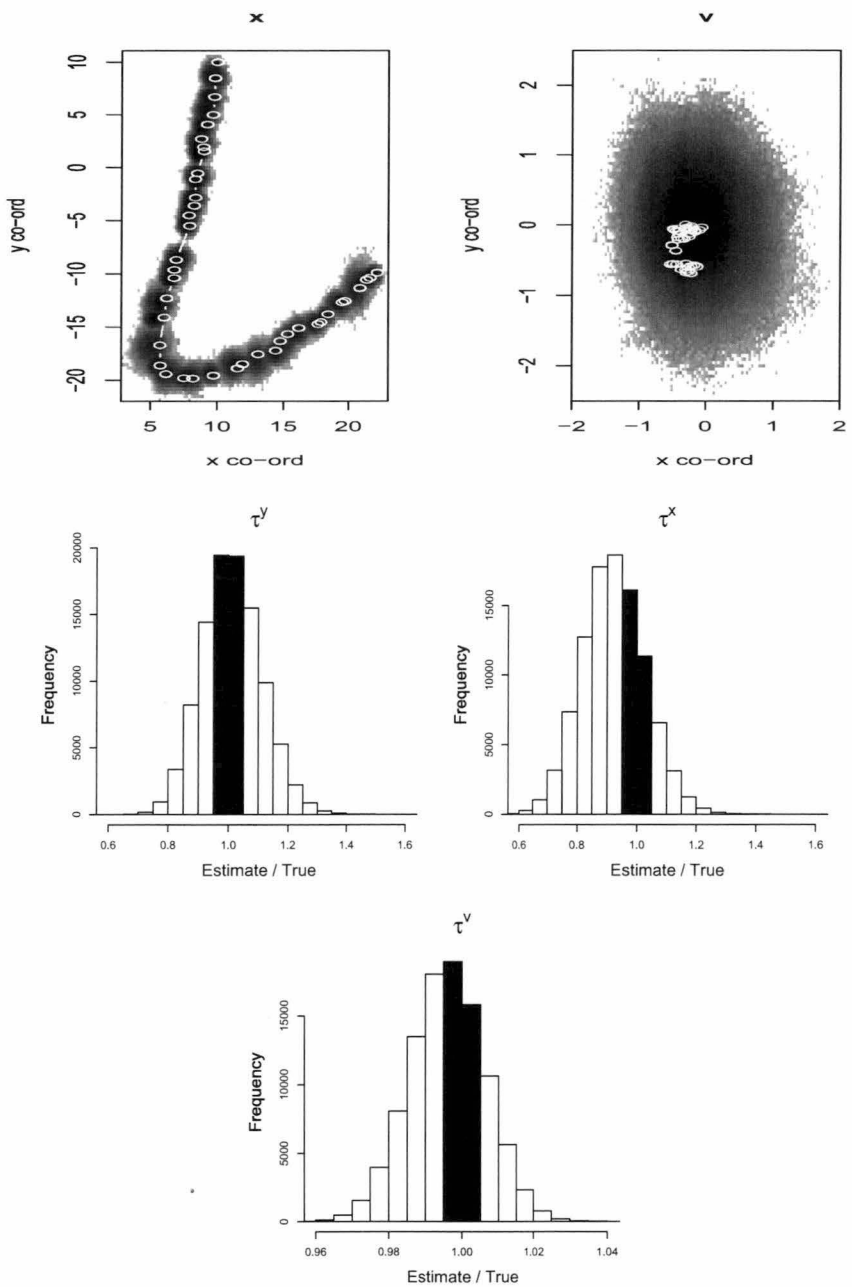


Figure 4.5: Estimation results for scenario 4, which uses an S-shaped declining adherence to a southerly advection field combined with an S-shaped increasing adherence to a gyre. The x and v plots show the true values as white, open circles and the 100 000 MCMC draws as a density plot. The precision parameters τ^y, τ^x and τ^v are shown as a fraction of their true values, so that a correct value is given as 1; the bars that span this value are shaded in black.

The reason for this is not clear. The precision for the observations appears unbiased.

Because the priors for the precision parameters are centred on their simulation values, their marginal posteriors are unbiased . These parameters are notoriously difficult to estimate (Rivot *et al.* , 2004) and, had the priors not been centred on the correct values, the results would have been biased, as demonstrated in Chapter 3. The precision parameters for the path τ^x and for the observations τ^y are estimated within a range of approximately 0.6 to 1.4 times the true value. A precise value was used for inertia τ^v so that the marginal for this parameter has less room for variation, ranging within 4% of the true value (Figures 4.2 to 4.5).

The precision matrix τ^B was not used to simulate the values of the coefficients B so there is no true value with which to compare the results. However, in the simulation, the coefficients for the two advection fields change in opposite directions, therefore the estimated covariate terms should be negative, as they are (Table 4.4). Greater variation is seen in the coefficients for the first advection field (0.65, 0.75, 0.73 and 0.64 in Table 4.4), compared with the second advection field (0.18, 0.30, 0.21 and 0.34) in all scenarios. This indicates greater variability in the B values, over the time steps, for the coefficients of the first advection field. The reason for this is not clear but it might also be indicated in Figure 4.6 where the median (the white line) for the estimates of the first advection field (left-hand plots) has a tendency to veer away from the overall trend during the final time steps.

Table 4.4: Average estimated (co)variance $(\tau^B)^{-1}$. For every one of the 100 000 posterior draws, the covariance between the vectors B_1 and B_2 was estimated, and the average over these 100 000 covariances is given in this table. Results are given for the four pairs of advections that represent the four simulation scenarios listed in Table 4.1.

Scenario 1		Scenario 2	
0.65	-0.06	0.75	-0.01
-0.06	0.18	-0.01	0.30

Scenario 3		Scenario 4	
0.73	-0.08	0.64	-0.06
-0.08	0.21	-0.06	0.34

While the true advection coefficients B_k range from 0 to 2, the 95% credibility interval for their estimates ranges much more widely, from roughly -2.5 to 3.5 (Figure 4.6). The negative values are of most concern because this causes the animal to

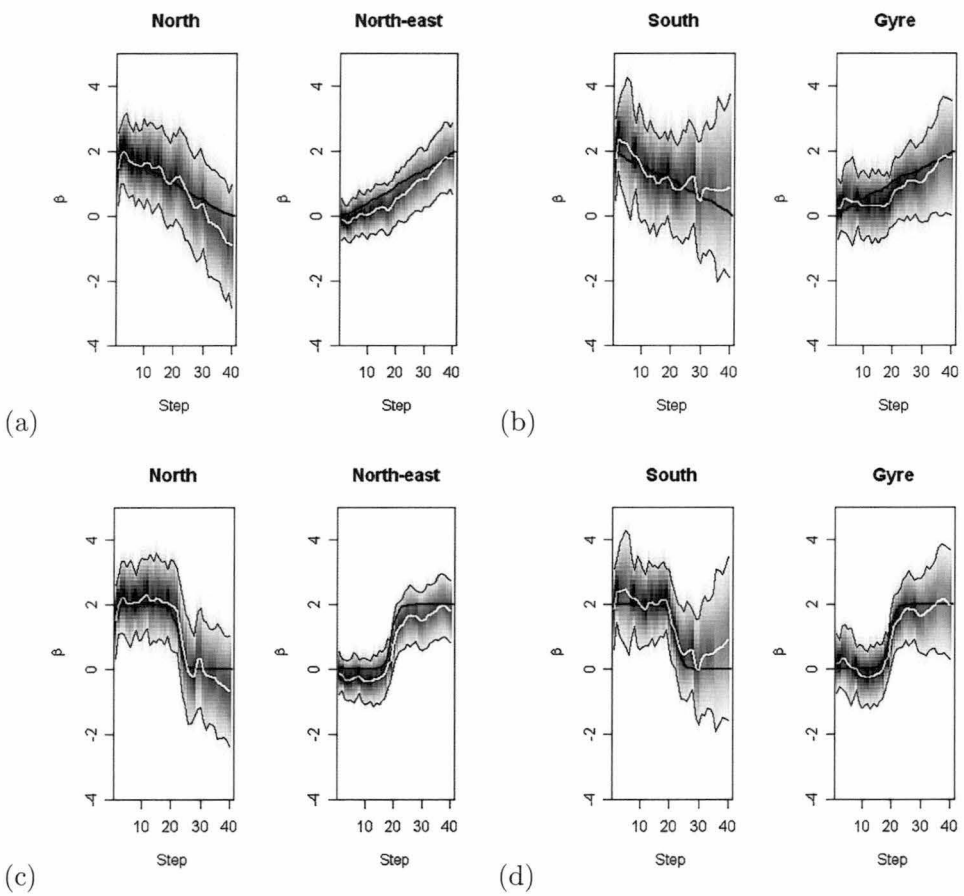


Figure 4.6: Marginal posteriors for the coefficients B for advection scenarios 1-4, listed in Table 4.1. The posterior median is indicated by a white line, a 95% credibility interval by thin black lines, and shading indicates posterior density. The true values used in the simulation are indicated by a thick black line.

move against the advection field instead of moving along with it. This results from confounding when the two advection fields flow in opposite directions. The model cannot distinguish between an animal that moves with an advection field that flows in a particular direction, and one that moves against an advection field that flows in the opposite direction. This results in a wide spread of B estimates of which some are negative.

Confounding also occurs when the two advection fields flow in the same direction. For example, in both scenarios 1 and 3 (which use the same advection fields - a northerly and a northeasterly flow) the coefficient B for the constant northerly flow is consistently overestimated while that for the constant northeasterly flow is underestimated, during the first 25 steps (Figure 4.6a and c). After the first 25 steps, the coefficient for northerly flow becomes negative, indicating southerly movement, and this is compensated by an increase in the coefficient for northeasterly flow.

It is difficult to distinguish advection from inertia. For example, the early steps of scenario 3, which should be following a northerly advection, are deviated to the west (Figure 4.6c). This is due to the chance drawing of a large, southwesterly, random deviate for inertia at the second step (ζ_2). The autocorrelation in inertia propagates that deviation so that the median values for the estimates of B for this part of the track are negative for the northeasterly flow, thus giving a southwesterly component to the track. The estimates of B for the northerly flow are correct (Figure 4.1c).

As the influence that an advection field has on movement wanes, so the model's ability to estimate its value declines. In all scenarios shown in Figure 4.6 the left-hand plot shows the advection field that dominates at the beginning of the walk and declines to zero by the end. By the final 10 steps the posterior median for the left-hand plots is biased by 0.5 to 1.0 units, whereas, the median for the corresponding right-hand plot shows little or no bias. This indicates that the two coefficients are not showing confounding with one another, rather, inertia is compensating for the mis-specification of the left-hand coefficients.

The model also has difficulty estimating the first coefficient value B_0, underestimating this in all cases (Figure 4.6). The prior for B_0 has a mean of zero, and the data lack sufficient information on this parameter to wholly overcome the effect of the prior.

The results for the gyre (Figure 4.6b and d, right-hand plots) show less bias than

those for the constant northeasterly flow, consistent with the finding in Chapter 3 that the coefficient for an advection field that constantly changes direction is more accurately estimated because there is less scope for confounding with inertia.

The shape of the functional form for B, whether linear or S-shaped, does not seem to effect the model's ability to correctly estimate the values of the coefficients.

4.5 Discussion

The estimation model was able to discern an increasing adherence to one advection field in conjunction with a decreasing adherence to another. The negative correlation between the sets of coefficients B for the two fields was correctly estimated. However, the values for the coefficients at the ends of the track (the earliest and latest time steps) are less well estimated than those in the middle, because the data are less informative about these. A random walk was used to describe the change in the coefficient values, therefore the value of the coefficient in any time step is constrained by those in the previous and next time steps. Those at the ends are thus less constrained.

A functional form could be imposed on the change in the coefficients. This would result in the coefficients at the ends of the track being estimated with greater precision. However, greater bias would also result unless this functional form closely approximated the true behaviour of the animal. In the (likely) absence of such knowledge it seems best to allow the coefficients to change according to a random walk, allowing the data to reveal the pattern of the change. The smoothness of the change is also a model estimated property, given by the precision for the errors in the coefficients τ^B. The change can be smooth (large precision values) or large fluctuations can be permitted between time steps (small precision values).

Confounding can occur between the coefficient estimates B when the advection fields used in the model flow in a similar direction in the area through which the animal moves. This was also seen in the model whose coefficients were time invariant (Chapter 3). Although the model is unable to discriminate the absolute values of these coefficients, their sum is expected to be correct, unless confounding with inertia occurs in addition. Unless prior information is available regarding the values of one of the coefficients, or their ratio, it will not be possible to discriminate such advection fields.

The model's ability to estimate negative B values can lead to confounding of advection fields that flow in opposite directions. This could be solved by disallowing negative B values - by using a truncated prior, or truncated likelihood.

Confounding with inertia was seen here, and also in the model that had time invariant (Chapter 3) coefficients. This problem is less severe when advection fields are used that change direction frequently. The model's inability to distinguish between persistence of direction due to inertia and that due to advection might be solved by combining the tracks from several animals, all following the same advection fields but having different random deviates, in a hierarchical model. The tracks simulated in these chapters had relatively few observed locations ($n = 21$), which is likely to have exacerbated the problem. Real applications are likely to use more observations, and more directionally variable advection fields.

The slight bias in the precision parameters for the path and the velocity requires further investigation. There may be confounding with the precision for the change in the coefficients τ^B, which could be investigated by simulating the change in B using specified τ^B values. Confounding could not be investigated in the simulation presented here because no τ^B value was used in the simulation.

A small value for n_0 yields a diffuse prior distribution but the distribution will be improper unless $n_0 \geq m$ (Gelman *et al.* , 2004). Because the Bayes factor is used to compare alternative models (Chapter 5), improper priors cannot be used Kass & Raftery (1995). It might be possible to compare models that have similar structure such that the improper priors, in a sense, cancel one another out (Kass & Raftery, 1995). However, dissimilar models cannot be compared. One of the models from Chapter 3 could not be compared with the model presented in this chapter if the Wishart prior was improper. Therefore, $n_0 = m$ was used to reflect uncertainty while still allowing the use of Bayes factors.

This chapter examines gradual replacement of one movement behaviour by another; no scenario is examined in which a single advection field is used, having a time variant co-efficient. Such a scenario would represent the weakening of some behaviour in favour of purely random motion (or the strengthening of some behaviour, away from purely random motion). The purpose of this thesis to develop methods for examining movement behaviour, it does not consider cases of purely random movement, unmodified by deliberate behavioural directives.

Chapter 5

Bayesian model comparison for movement model

5.1 Abstract

This chapter examines methods of model comparison suitable for application to the modelling framework presented in Chapter 3. Two methods are applied - Bayes factor, and the deviance information criterion DIC. The Bayes factor, the primary means of Bayesian model comparison, requires the estimation of a difficult integral - the marginal likelihood. A number of estimators are available for this quantity - this chapter uses direct Monte Carlo estimation, the harmonic mean estimator, and three variants on Chib's Candidate's estimator. The details of the application of Chib's method are dependent on the type MCMC sampling performed. Applications are performed for Gibbs sampling, Metropolis Hastings within Gibbs, and one-block Metropolis-Hastings sampling.

A model of movement behaviour, drawn from the family of models presented in Chapter 3, was used to simulate a track for an animal moving northwards along a preferred line of bathymetry. An advection field relating to an alternative movement hypothesis was proposed, suggesting that the animal simply moved in a northerly direction. The same movement model was then applied to the simulated track, using either the 'true' advection field, or the alternative field.

However, the harmonic mean estimator gave different estimates of the Bayes factor to those produced by the Chib estimators and was unable to discriminate between the two proposed behaviours. The reasons for the failure of this published method are elucidated in Chapter 6.

Very similar estimates of Bayes factor resulted from the three Chib variants. When highly diffuse priors were not used, the Bayes factor successfully discriminated which of the two proposed advection fields was the one used to simulate the track. However, when a highly diffuse prior is used (variance=100) Bayes factor provides no evidence in favour of one model over the other.

The DIC is less sensitive than the Bayes factor to the choice of prior, providing evidence in favour of the correct model in all simulations. It is much easier to calculate than the Bayes factor, requiring only a set of draws from the posterior. The DIC is therefore the preferred estimator for model comparison using the movement modelling state space model presented in this thesis.

Chapter 3 showed that the estimates of the precision parameters for the path τ^x, and the observations τ^y, are strongly influenced by the choice of prior. When the prior mean for these parameters is incorrect, the estimates are also incorrect, having the same mean as the prior. Nevertheless, these estimates did not influence model comparison so that the correct advection field was selected regardless of the mean value of the prior for the precision parameters.

5.2 Introduction

Chapter 3 presented a model of movement behaviour and demonstrated its ability, in simulation, to recover the 'true' parameter values used to simulate the track of a moving animal. However, in these simulations, the advection fields used in the estimation model were similar to, or exactly the same as, those used in the simulation model. Except for different advection fields and changes in parameter values, the estimation and simulation models had the same structure. In reality, the complexity of movement behaviour is unlikely to be as closely replicated by the estimation model. For this reason, the exact parameter values estimated will not always be of great interest to the investigator, as they may not have close biological analogues. Instead, in such cases, it would be more reasonable to compare competing hypotheses regarding behaviour in order to evaluate which are most consistent with observed track data. In this way, it is hoped to eliminate hypothesized behaviours that are least likely to have given rise to the observed track. In this way proposed covariates of behaviour to which tracked animals do not appear to be responding can be eliminated.

The primary means of model comparison, in a Bayesian context, is the Bayes factor (Jeffreys, 1935, 1961), which can be interpreted as the posterior probability of one model compared with others (Carlin & Louis, 2000). Any number of models may be compared in this way and the Bayes factor can also be used as a weight for Bayesian model averaging (described in Section 2.7.1).

In order to calculate the Bayes factor, the marginal likelihood, which is the normalising constant for the posterior, must be calculated. This involves an integral of the likelihood, multiplied by the priors, across all possible values for the parameters and unknown states. Improper priors cannot be used in association with the Bayes factor because the unknown normalising factor of the improper prior will become incorporated into the marginal likelihood, preventing valid comparison between different models. Kass & Raftery (1995) notes that several authors, including Jeffreys, have used improper priors. However, these were cases where the two models being compared were very similar. It might be expected that the normalising constants of their priors are also similar and therefore cancel one another out in the ratio of their marginal likelihoods (which is the Bayes factor).

Although MCMC provides accurate methods for estimating the shape of the posterior, the normalising factor remains unknown. However, the set of draws from the posterior that MCMC does provide can be used to estimate the marginal likelihood. Many such estimators have been proposed. Path sampling (Meng & Wong, 1996) and its special case, bridge sampling (Gelman & Meng, 1998), are among the most powerful but are complex in implementation and rely on good choices of importance distributions. Reciprocal importance sampling, RIS (Gelfand & Dey, 1994) is, in turn, a special case of bridge sampling. This is relatively simple to implement, but again requires careful selection of a distribution. This must be narrower-tailed than the posterior for all of the model's parameter space; a difficult achievement when dealing with a high dimensional problem. A special case, in turn, of RIS, is the harmonic mean estimator (Newton & Raftery, 1994). This method is very simple to use, requiring only a set of draws from the posterior. Although it is known to be unstable, it is said to be sufficiently accurate for practical model comparison (notably Rosenkrantz & Raftery (1994); Carlin & Chib (1995); Kass & Raftery (1995); Raftery (1996); Sorensen & Gianola (2002)). The accuracy of the harmonic mean estimator is revisited in Chapter 6, where it is shown to be unsuitable for practical model comparison.

Laplace's method (Tierney & Kaldane, 1986) is a non-MCMC based approach that approximates the posterior using a multivariate normal distribution. The method

is straightforward to code, but is unreliable in high dimensional problems, such as the one presented here. Furthermore, it requires the inversion of a Hessian matrix, which in this case was large enough to generate computational problems.

The marginal likelihood may be estimated by direct Monte Carlo integration by making draws from the prior and taking the ergodic average of the corresponding posterior densities. This method is trivial to code, and is reliable when the priors are precise, but as shown in Chapter 6, when the prior is diffuse, it converges too slowly to be of practical use.

Chib's Candidates Method (Chib, 1995; Chib & Jeliazkov, 2001) is complex to use in that it requires the calculation of conditional distributions or the selection of appropriate approximations to these. However, when MCMC sampling has been used these will already be available. This method is accurate even in high dimensional problems. However, it has high computational demands (although still much lower than direct Monte Carlo integration). As well as using a set of MCMC draws from the posterior, it also requires additional reduced runs. An attempt is made to reduce the computational burden by modifying the sampling approach, as described later.

The difficulties in calculating the Bayes factor have encouraged the use of several, very much more easily calculated, information criteria. These are the Bayesian information criterion BIC (or the Schwarz criterion) (Schwarz, 1978), Akaike's information criterion AIC (Akaike, 1973), and the deviance information criterion DIC (Spiegelhalter *et al.* , 1998, 2002). Comprehensive discussion on these methods can be found in DiCiccio *et al.* (1997), Kadane & Lazar (2004), and Kass & Raftery (1995), and greater detail than that presented in this chapter is given in Section 2.7.1.

Provided the prior has comparable precision to the likelihood, it can be shown that AIC is asymptotically equivalent to the Bayes factor (Kadane & Lazar, 2004; Kass & Raftery, 1995). This implies that as the sample size increases, both the prior and the likelihood become more informative - an unlikely situation. AIC has been found to overestimate the number of parameters needed by the model so that BIC, which differs only in that it more heavily penalizes models with more parameters, is favoured in this respect (Kass & Raftery, 1995; Kadane & Lazar, 2004). The justification for using BIC is that it is closely related to AIC, without the bias towards less parsimonious models, which in turn is justified as being asymptotically equivalent to the Bayes factor provided the prior becomes more informative as the sample size increases. Apart from the problem of precise priors, asymptotic qualities

are unlikely to be of help in movement behaviour problems where large sample sizes are unlikely. AIC and BIC require that the sample size, and BIC requires that the number of model parameters, be known. These methods are therefore not useful for model comparison involving SSMs where these quantities are difficult to define. For a typical SSM, the calculation of the true likelihood requires a high dimensional integration, so that iterative estimation of the maximum likelihood is likely to be prohibitive. AIC and BIC are used because, although less rigorous than the Bayes factor, they are computationally relatively undemanding (Kass & Raftery, 1995). This advantage does not extend to the case of an SSM where the likelihood is difficult to obtain.

The DIC was developed as a solution to the problem of identifying the number of parameters, and the sample size, for hierarchical models. When the model is hierarchical, the sum of the number of parameters and the number of states will give an overestimate of the sample size, whereas the number of parameters, alone, will give an underestimate. The sample size, too, may be unclear, for example, when the data consist of multiple measures from several individuals (Carlin & Louis, 2000). Both of these problems pertain to the movement model presented here. In non-Bayesian terms, DIC can be regarded as a generalization of the AIC, but it has additional, Bayesian, justification through its use of the deviance statistic (Spiegelhalter *et al.* , 2002). DIC is easy to calculate from a set of MCMC draws from the posterior but has the disadvantage that different model parameterisations (leading to different marginalizations of the likelihood) can give different DIC values (Carlin & Louis, 2000).

The Bayes factor is known to be sensitive to the choice of prior, in particular, to diffuse priors (Kadane & Lazar, 2004; Aitkin, 1991; Gelfand, 1996) - a disadvantage not shared by AIC, BIC and DIC. Movement behaviour is, as yet, little understood so that it would be preferable to use a model comparison measure that is robust to the choice of prior.

This chapter presents a simulation of the track of an animal that is moving according to a set of advection fields, representing a specific movement behaviour. It is assumed that the behaviour that gave rise to this track is unknown, and several competing hypotheses regarding movement behaviour are tested to see which is given best support by the data. Each competing hypothesis leads to an alternative set of advection fields, and therefore to an alternative version of the movement model presented in Chapter 3. For each of these models, the marginal likelihood is calculated, by several methods, along with the DIC, to evaluate whether both the

Bayes factor and DIC may be used to identify which model, and therefore which hypothesis, is best supported by the observed track.

5.3 Methods

5.3.1 Bayes factor

The Bayes factor is interpreted on a scale originally proposed by Jeffreys (1961). Of the many variants available, the more conservative scale put forward by Kass & Raftery (1995), Table 5.1, is used here.

Table 5.1: Scale presented by Kass and Raftery (1995) after Jeffreys (1961) for interpretation of the Bayes factor (B_{12}) for model H_1 compared with model H_2.

$\log(B_{12})$	Evidence for H_1
<1	Not worth more than a bare mention
1 to 3	Positive
3 to 5	Strong
> 5	Very strong

Given two models, M_1 and M_2, each with equal prior model probability, the Bayes factor B_{12} is the posterior odds of model M_1 compared with model M_2. It can be calculated as the ratio of the marginal likelihoods (Kass, 1993)

$$B_{12} = \frac{m(y \mid M_1)}{m(y \mid M_2)} \tag{5.1}$$

where

$$m(y \mid M_i) = \int p(y \mid \theta, M_i)\, p(\theta \mid M_i)\, d\theta.$$

Here $p(y \mid \theta,\ M_i)$ is the true likelihood (see Section 2.5.2) of the data y given parameters θ and model M_i, and $p(\theta \mid M_i)$ is the prior.

Thus the calculation of the Bayes factor reduces to the calculation of marginal likelihoods $m(y)$. Suppressing model dependence

$$m(y) = \int p(y \mid \theta)\, p(\theta)\, d\theta. \tag{5.2}$$

In the case of the hierarchical model presented in this thesis (introduced in Chapter 3), equation 5.2 includes unobserved states (equation 3.7), so that the marginal likelihood is

$$m(y) = \int p(y \mid x, v, \theta, V)\, p(x \mid v, \theta, V)\, p(v \mid \theta, V)\, p(\theta)\, dx\, dv\, d\theta. \tag{5.3}$$

The parameter values, denoted θ, are the precision terms τ^x, τ^v, and τ^y, as well as the advection coefficients β. The true locations x, and the velocities v, are states. The control variable V, which is not dependent on the states, is the totality of all advection data. From V, a subsample $\mathcal{X}$ is drawn, representing the advections to which the moving individual responded. This is dependent on x, $\mathcal{X} = \mathcal{X}(x)$, unless advection is constant at all locations.

From equation 5.3 it can be seen that the marginal likelihood can either be regarded as the expectation of $p(y \mid x, v, \theta, V)$ with respect to the priors for all the states $p(x \mid v, \theta, V)$ and $p(v \mid \theta, V)$ and parameters $p(\theta)$, or as the expectation of the complete data likelihood $p(y \mid x, v, \theta, V)\, p(x \mid v, \theta, V)\, p(v \mid \theta, V)$ with respect to the prior $p(\theta)$. In the description below, the first interpretation is used, so that the 'likelihood' is $p(y \mid x, v, \theta, V)$. For clarity, in this thesis this is termed the *full likelihood*. Section 2.5.2 gives a more detailed discussion on likelihoods and their terminology.

5.3.2 Monte Carlo estimation

The marginal likelihood (equation 5.3) may be estimated by direct Monte Carlo integration by making draws from the prior and taking the ergodic average of the corresponding unnormalized posterior densities

$$\hat{m}(y) = \frac{1}{J} \sum_{j=1}^{J} p(y \mid x^{(j)}, v^{(j)}, \theta^{(j)}, \mathcal{X}^{(j)}). \tag{5.4}$$

Here the superscript (j) indicates the j'th of J draws of the parameters θ from

their priors (equations 3.13 and 3.14), and of the states x from $p(x \mid v, \theta, V)$ and v from $p(v \mid \theta, V)$. Drawing x involves drawing $x_{i,1}$, $i = 1, 2$ from equation 3.11 and then recursively applying equation 3.3 using values of $\xi_{i,k}$ drawn from equation 3.5. Drawing v, similarly, involves drawing $v_{i,1}$ from equation 3.12 and recursively applying equation 3.4 using draws of $\zeta_{i,k}$ from equation 3.6. The advections $\mathcal{X}^{(j)}$ are calculated, using the relevant advection co-model, to be those operating at the locations $x^{(j)}$.

5.3.3 Harmonic Mean

Given G MCMC draws of the parameters and states from the posterior, the harmonic mean estimate of the marginal likelihood (Newton & Raftery, 1994) is

$$\hat{m}(y) = \left[\frac{1}{G} \sum_{g=1}^{G} p(y \mid x^{(g)}, v^{(g)}, \theta^{(g)}, \mathcal{X}^{(g)})^{-1} \right]^{-1}. \tag{5.5}$$

A derivation for this equation is given in Chapter 6.

5.3.4 Chib's Candidates Method

The precise details of Chib's Candidates method depend on the form of the MCMC sampling routine used. The following section describes three variants implemented in this thesis. First, Chib's method for Gibbs sampling uses the conditional distributions for the parameters and states to calculate ordinates for each using Rao-Blackwellization (see Robert & Casella, 1999). This version is computationally intensive, requiring that the MCMC procedure be executed several times. Second, Chib's method for single block Metropolis-Hastings sampling requires only a single run of the MCMC routine. However, this MCMC variant is less efficient because all parameters and states are sampled in a single Metropolis-Hastings step even though conditional distributions are available for all but (possibly) one state (x). Third, Chib's method for multiple block Metropolis-Hastings sampling is presented. When conditional distributions are available for all parameter blocks, this reduces to the Gibbs sampling variant. The calculations performed for this thesis use a combination of Gibbs and Metropolis-Hastings sampling. This makes use of the Chib method for multiple block Metropolis-Hastings sampling but some of the calculations reduce to the simpler Gibb's sampling form. As this methodology is rather

involved, Section 5.3.4, page 116, describes the procedure in point form.

The Chib's Candidate's method (Chib, 1995; Chib & Jeliazkov, 2001) estimator recognizes that Bayes theorem holds for any specific set of parameter values so that, rearranging Bayes theorem (and using ψ to indicate all parameters and states)

$$m(y) = \frac{p(y \mid \psi^*)\, p(\psi^*)}{p(\psi^* \mid y)} \tag{5.6}$$

where $*$ indicates a specified set of values. For greatest efficiency Chib (1995) recommends using values corresponding to a high density area of the posterior. The posterior draw yielding the greatest full likelihood value is used here. All terms in the numerator on the right hand side of this equation may be calculated exactly, but $p(\psi^* \mid y)$ is more difficult.

For Gibbs sampling

Chib (1995) presents a method of estimating $p(\psi^* \mid y)$ from a set of draws from the posterior, made using Gibbs sampling. First, $p(\psi^* \mid y)$ is factorized into several ordinates, each of which is estimated separately. If MCMC is conducted using B parameter and state blocks $\psi_1, ..., \psi_B$ then the factorisation is

$$p(\psi^* \mid y) = p(\psi_1^* \mid y)\, p(\psi_2^* \mid y, \psi_1^*) \ldots p(\psi_B^* \mid y, \psi_1^*, \ldots \psi_{B-1}^*) \tag{5.7}$$

The first ordinate is estimated using the results of the full MCMC run. For each draw from the posterior (yielding $\psi_1^{(g)} \ldots \psi_B^{(g)}$, where for all draws $g = 1, \ldots G$) the value of ψ_1 is set equal to the chosen value ψ_1^* and the full conditional distribution for ψ_1 is used to yield the estimate

$$\hat{p}(\psi^* \mid y) = G^{-1} \sum_{g=1}^{G} p\left(\psi_1^* \mid \psi_2^{(g)}, \ldots \psi_B^{(g)}, y\right). \tag{5.8}$$

The second ordinate is estimated by fixing the value of ψ_1 at ψ_1^*, performing a reduced MCMC run that draws values for all parameters except ψ_1, replacing the resulting ψ_2 draws with ψ_2^* and then averaging over the conditional for ψ_2. Because

the value of ψ_1 was fixed at ψ_1^* prior to performing a reduced run, the resulting draws will be from $p\left(\psi_2^* \mid \psi_1^*, \psi_3, \ldots \psi_B, y\right)$.

Thus for each ordinate after the first, an additional run of the MCMC algorithm is required, each time fixing another parameter block at its specified $*$ values. Unless there are latent data, the final ordinate requires no additional MCMC run - simply fix all blocks at their chosen values and calculate the corresponding density of the conditional for ψ_B^*. Although this model includes latent locations, these are states and because this is a Bayesian implementation and all states and parameters are random variables, these locations are treated in the same way as parameters of the model. Therefore, no additional MCMC run is required.

For single block Metropolis-Hastings sampling

When the full conditional distribution is unavailable (such as that for the states x when the advection field is not constant) then neither Gibbs sampling nor equation 5.8 can be used. Chib & Jeliazkov (2001) present an alternative formula for estimating the marginal likelihood using a set of draws made by Metropolis-Hastings sampling. Consider Metropolis-Hastings sampling from the posterior using proposal distribution $q(\psi, \psi' \mid y)$ for a move from ψ to ψ'. The probability of accepting the proposed value is

$$\alpha(\psi, \psi' \mid y) = \min\left\{1, \frac{p(y \mid \psi')\, p(\psi')\, q(\psi, \psi' \mid y)}{p(y \mid \psi)\, p(\psi)\, q(\psi', \psi \mid y)}\right\}. \tag{5.9}$$

By the reversibility of the subkernel (Gilks *et al.* , 1996; Chib & Jeliazkov, 2001)

$$\alpha(\psi, \psi^* \mid y)\, q(\psi, \psi^* \mid y)\, p(\psi \mid y) = \alpha(\psi^*, \psi \mid y)\, q(\psi^*, \psi \mid y)\, p(\psi^* \mid y). \tag{5.10}$$

Integrating both sides with respect to ψ, and making $p(\psi^* \mid y)$ the subject of the formula

$$p(\psi^* \mid y) = \frac{\int \alpha(\psi, \psi^* \mid y)\, q(\psi, \psi^* \mid y)\, p(\psi \mid y)\, d\psi}{\int \alpha(\psi^*, \psi \mid y)\, q(\psi^*, \psi \mid y)\, d\psi}. \tag{5.11}$$

The numerator can be regarded as an expectation with respect to $p(\psi \mid y)$ (the posterior), and the denominator as an expectation with respect to $q(\psi^*, \psi \mid y)$. Therefore, a Monte Carlo approximation for $p(\psi^* \mid y)$ is given by

$$\hat{p}(\psi^* \mid y) = \frac{G^{-1} \sum_{g=1}^{G} \alpha(\psi^{(g)}, \psi^* \mid y)\, q(\psi^{(g)}, \psi^* \mid y)}{J^{-1} \sum_{j=1}^{J} \alpha(\psi^*, \psi^{(j)} \mid y)} \tag{5.12}$$

where $\psi^{(g)}$, $g = 1 \ldots G$ are draws from the posterior and $\psi^{(j)}$, $j = 1 \ldots J$ are draws from the proposal distribution $q(\psi^*, \psi \mid y)$. Each time an MCMC run, or reduced run (which are still required if sampling is performed in several blocks), is performed, an additional set of draws is made from the proposal distribution.

In the implementation presented in this thesis, the conditional distributions are available for all parameters and states except, sometimes, for x, so that a Metropolis-Hastings step is used for x. When the advection field is constant, X is the same at all locations and the proposal distribution used for x reverts to the full conditional distribution for x. In this case, the Metropolis-Hastings step reduces to Gibbs sampling (shown in Chapter 3, equation 3.28 and accompanying text). Note that equation 5.12 will likewise reduce to equation 5.8 because the proposed value will always be accepted $\alpha(\cdot) = 1$, and $q(\cdot)$ in equation 5.12 will become the full conditional, $p(\cdot)$ in equation 5.8.

For multiple block Metropolis-Hastings within Gibbs sampling

The previous sections described Chib's method for Gibbs sampling and for single-block Metropolis-Hastings sampling. This section does not introduce any new formulae, but shows how the techniques of the previous sections were applied to the Metropolis-Hastings within Gibbs sampling routine used in this thesis. The method is presented in terms of the parameters and states of this model. Note that when the model uses a constant advection field, the proposal distribution used for the path x becomes the full conditional for x so that the Metropolis-Hastings step used to draw x reduces to a Gibbs sampling step and the method described in this section reduces to Chib's method for Gibbs sampling (page 112). This procedure is described again, in point form, on page 116.

The parameters (and states) were sampled as four blocks using a mixture of Metropolis-

Hastings and Gibbs sampling. Note that the precision parameters τ^y, τ^x and τ^v are independent so that all three may be drawn together as a single τ block. The four blocks were x, v, τ and β so that Chib's equation 5.7 becomes

$$\begin{aligned} p(\psi^* \mid y) &= p(x^*, v^*, \tau^*, \beta^* \mid y) \\ &= p(x^* \mid y)\, p(v^* \mid y, x^*)\, p(\beta^* \mid y, x^*, v^*)\, p(\tau^* \mid y, x^*, v^*, \beta^*) \end{aligned} \quad (5.13)$$

The first, full, run of the sampler yielded $G = 100\,000$ draws of $x^{(g)}$, $v^{(g)}$, $\tau^{(g)}$ and $\beta^{(g)}$, $g = 1 \ldots G$ with associated full likelihood values. Gibbs sampling was used for v, τ and β but a Metropolis-Hastings step was used for x, as described in Chapter 3, Section 3.3.6, using the 'conditional' for x (equation 3.24) as the proposal distribution. The advection values X used in this equation were those corresponding with the values of x from the previous draw (or the starting values).

From these G draws, and their associated full likelihoods, the set of values x^*, v^*, τ^* and β^* was selected that gave the largest likelihood.

Having performed the Metropolis-Hastings step for x first, only the more efficiently Gibbs sampled states and parameters are left for the two subsequent reduced runs, in which x is held fixed at x^*. First, with x fixed at x^*, a further $J = 100\,000$ values were drawn of $v^{(j)}$, $\beta^{(j)}$, and $\tau^{(j)}$, $j = 1 \ldots J$ using Gibbs sampling. In addition, $x^{(j)}$ was drawn from the proposal distribution $q(x^*, x)$. Here the advection values X used were those corresponding with the locations x^*, $X(x^*)$.

When performing the Metropolis-Hastings step for x, the transition probability $\alpha(a, b \mid y)$ for stepping from $x = a$ to $x = b$, is given by equation 3.25 and subsequent explanation in Chapter 3. The stored values drawn during the full run provide the numerator, and the values from the reduced run give the denominator of equation 5.12, which is used to calculate an estimate $\hat{p}(x^* \mid y)$ of the first ordinate of equation 5.13

$$\hat{p}(x^* \mid y) = \frac{G^{-1} \sum_{g=1}^{G} \alpha(x^{(g)}, x^* \mid y)\, q(x^{(g)}, x^* \mid y)}{J^{-1} \sum_{j=1}^{J} \alpha(x^*, x^{(j)} \mid y)}. \quad (5.14)$$

As described above, if the advection forces used in the movement model are constant at all locations so that X is not a function of x, then equation 5.14 reduces to an

ergotic average of the conditional for x, of the same form as equation 5.15.

The second ordinate of equation 5.13 is given by replacing all of the draws $v^{(j)}$ by v^* and then using the remaining draws $\beta^{(j)}$ and $\tau^{(j)}$ to calculate the corresponding density of the conditional distribution for v, $p(v^* \mid y, x^*, \tau^{(j)}, \beta^{(j)})$ (equation 3.21), and taking their ergotic average

$$\hat{p}(v^* \mid x^*, y) = J^{-1} \sum_{t=1}^{J} p(v^* \mid y, x^*, \tau^{(j)}, \beta^{(j)}). \tag{5.15}$$

The third ordinate in equation 5.13, $p(\beta^* \mid y, x^*, v^*)$ is calculated similarly to the second by fixing v at v^*, performing a second reduced Gibbs sampling run which yields $\beta^{(t)}$ and $\tau^{(t)}$, $t = 1 \ldots, T$ and then replacing the $\beta^{(t)}$ by β^* and taking the average of the corresponding densities of the conditional distribution for β (equation 3.22)

$$\hat{p}(\beta^* \mid x^*, v^*, y) = T^{-1} \sum_{t=1}^{T} p(\beta^* \mid y, x^*, v^*, \tau^{(t)}). \tag{5.16}$$

The fourth, and final, ordinate $p(\tau^* \mid y, x^*, v^*, \beta^*)$ requires no further MCMC sampling, because it is simply the density of the conditional distribution for τ when all states and parameters are set equal to their high density values, ψ^*.

Substitution of the four ordinates into equation 5.13 gives an estimate of the posterior at a high density location $p(\psi^* \mid y)$. Substituting that, in turn, into the numerator of equation 5.6, along with the prior $p(\psi^*)$ density and the full likelihood $p(y \mid \psi^*)$ at this point, provides an estimate of the marginal likelihood. Note that the product of the prior and the likelihood is equal to the unnormalized posterior so that equation 3.7 evaluated at ψ^* provides the numerator for equation 5.13.

Algorithm for Metropolis-Hastings within Gibbs MCMC and Chib's method

Applying Chib's method for calculating the marginal likelihood to Metropolis-Hastings within Gibbs sampling requires a full run of the MCMC sampler, followed by several reduced runs. At each stage, one or more quantities are calculated and stored for

later substitution into Chib's equations. This procedure was described above but, being relatively involved, is presented again in point form.

First, the Metropolis-Hastings algorithm is given.

1. Choose starting values for x, v, β, and τ: $x^{(0)}$, $v^{(0)}$, $\beta^{(0)}$, and $\tau^{(0)}$. Those for $x^{(0)}$ were given by the observations y with a cubic spline used to interpolate the latent locations. Those for $v^{(0)}$ were given by the first differences of $x^{(0)}$.

2. Draw a, candidate values for x, from $p(x \mid X(x^{(0)}), v^{(0)}, \beta^{(0)}, \tau^{(0)})$. Evaluate $\alpha(\psi, \psi')$ where ψ is ($x^{(0)}$, $v^{(0)}$, $\beta^{(0)}$, $\tau^{(0)}$) and ψ' is (a, $v^{(0)}$, $\beta^{(0)}$, $\tau^{(0)}$). Set $x^{(1)} = a$ with probability $\min(1,\ \alpha(\psi, \psi'))$, else set $x^{(1)} = x^{(0)}$.

3. Draw $v^{(1)}$ from $p(v \mid x^{(1)}, \beta^{(0)}, \tau^{(0)})$.

4. Draw $\beta^{(1)}$ from $p(\beta \mid x^{(1)}, v^{(1)}, \tau^{(0)})$.

5. Draw $\tau^{(1)}$ from $p(\tau \mid x^{(1)}, v^{(1)}, \beta^{(1)})$.

6. Repeat from (b) yielding $x^{(g)}$, $v^{(g)}$, $\beta^{(g)}$, and $\tau^{(g)}$, $g = 1, \ldots, G$, where G is 100 000.

Chib's method for estimating the marginal likelihood from a set of draws made using the Metropolis-Hastings within Gibbs algorithm above, follows.

1. Execute the full MCMC algorithm (given above). From the resulting G sets of draws and their corresponding likelihood values, select those that correspond to the greatest likelihood: x^*, v^*, β^*, τ^*.

   (a) Calculate the numerator of equation 5.14, $S_1 = G^{-1} \sum_{g=1}^{G} \alpha(x^{(g)}, x^* \mid y)\ q(x^{(g)}, x^* \mid y)$.

2. Fix $x = x^*$, and set the starting values $v^{(0)} = v^*$, $\beta^{(0)} = \beta^*$, $\tau^{(0)} = \tau^*$. Repeat steps (c) to (e) (with x always equal to x^*) above yielding $v^{(j)}$, $\beta^{(j)}$, and $\tau^{(j)}$, $j = 1, \ldots, J$, where J is 100 000.

   (a) In addition, for each loop through (c) to (e) above, draw $x^{(j)}$ from $p(x \mid X(x^*), v^{(j)}, \beta^{(j)}, \tau^{(j)})$.

   (b) Calculate the denominator of equation 5.14, $S_2 = J^{-1} \sum_{j=1}^{J} \alpha(x^*, x^{(j)} \mid y)$.

   (c) Estimate the ordinate for x using equation 5.14, $\hat{p}(x^* \mid y) = S_1/S_2$.

   (d) Estimate the ordinate for v, $\hat{p}(v^* \mid x^*, y)$, using equation 5.15.

3. Fix $x = x^*$, $v = v^*$ and set the starting values $\beta^{(0)} = \beta^*$, and $\tau^{(0)} = \tau^*$. Repeat steps (d) to (e) above yielding $\beta^{(t)}$ and $\tau^{(t)}$, $t = 1, \ldots, T$, where T is 100 000.

   (a) Estimate the ordinate $\hat{p}(\beta^* \mid x^*, v^*, y)$ using equation 5.16.

4. Estimate the ordinate $\hat{p}(\tau^* \mid y, x^*, v^*, \beta^*)$ by evaluating the conditional distribution for τ at the starred values.

5. Estimate the marginal likelihood by evaluating the likelihood multiplied by the prior, at the starred values, and divide by the product of the four ordinates, equation 5.6.

5.3.5 DIC

The deviance statistic $D(y, \theta)$ for a model for data y given parameters θ is often defined as a function of its likelihood $p(y \mid \theta)$ (e.g. Sorensen & Gianola, 2002; Gelman *et al.* , 2004) (discussion on which likelihood this is, follows)

$$D(y, \theta) = -2 \log p(y \mid \theta) \tag{5.17}$$

although in truth, deviance also includes a term that relates to the likelihood of the saturated model; a term that can be considered to be a function of the data alone, not of the model (Carlin & Louis, 2000; Pawitan, 2001). For this reason, it can be ignored in the context of model comparison between different models that nevertheless use the same data. The deviance can be used to indicate lack-of-fit (Carlin & Louis, 2000).

Using equation 5.17 as the definition of deviance, Spiegelhalter *et al.* (1998, 2002)'s deviance information criterion DIC can be expressed as the sum of the expected deviance with respect to the posterior $\overline{D}$ and the expected number of parameters p^D

$$DIC = \overline{D} + p^D. \tag{5.18}$$

DIC is thus a combination of a measure of lack-of-fit $\overline{D}$, and a measure of model complexity p^D, so that the preferred model is the one that returns the smallest DIC.

The concept of the expected number of parameters p^D, recognizes that hierarchical models, where hyperpriors introduce a degree of dependence between parameters, have lower dimensionality than they would, had the hyperprior not been specified, but that the degree of this reduction would depend somewhat on the information content of the data. Spiegelhalter *et al.* (2002) suggests that an estimate of the effective number of parameters may be given by the difference between the posterior expectation for the deviance $\overline{D}$, and the deviance evaluated at the posterior expectation of the parameters $D(\overline{\theta})$

$$p^D = \overline{D} - D(\overline{\theta}). \tag{5.19}$$

Combining equations 5.18 and 5.19

$$DIC = 2\overline{D} - D(\overline{\theta}). \tag{5.20}$$

Given the set of draws $\theta^{(g)}$, $g = 1 \ldots G$ from the posterior that MCMC provides, a Monte Carlo estimate for the expectation $\overline{D}$ with respect to the posterior, $\hat{\overline{D}}$, is

given by the average of the log likelihood evaluated at each of the draws (Sorensen & Gianola, 2002)

$$\hat{\overline{D}} = \frac{1}{G}\sum_{i=1}^{G} -2\log p(y \mid \theta^{(g)}). \tag{5.21}$$

The same set of draws can be used to estimate the posterior mean $\overline{\theta}$ by calculating the average, from the sample, of each component of the vector θ. Spiegelhalter *et al.* (2002) recommends using the geometric mean of the precision parameters (that is, the arithmetic mean of their logged values). Therefore, in this case, the geometric mean of τ^y, τ^x, and τ^v was used, whereas for all other quantities the arithmetric mean was used. Given the posterior mean $\overline{\theta}$, the deviance at the posterior mean $D(\overline{\theta})$ can be found by substitution of $\overline{\theta}$ into equation 5.17. In this way, the DIC is easily calculated given a set of MCMC draws from the posterior.

A difficulty when using DIC is specifying what the 'likelihood' is. Spiegelhalter *et al.* (2002) recommends using a likelihood from which nuisance parameters have been integrated out (the true likelihood) but acknowledges that this may not always be done for numerical reasons. The same model can give different DIC values depending on what marginalisation is used. Although it is not clear what marginalization to use for the likelihood, it is clear that the full likelihood would be an inappropriate choice because it ignores the dependence of the states on the parameters. Here, the complete data likelihood is used (see Section 2.5.2 for more information on likelihoods in state space models).

5.3.6 MCMC and convergence

MCMC was applied in two ways in this Chapter. First, MCMC was applied as Metropolis-Hastings within Gibbs, which reduces to Gibbs sampling when a constant advection is used so that the advection X is not a function of location x and the proposal distribution for x is the full conditional distribution for x.

Second, all parameters and states were drawn as a single block using Metropolis-Hastings sampling. This was an attempt to reduce the computational load of Chib's method by avoiding reduced runs. The proposal distribution was the product of the conditional distributions for all parameters and states (including the 'conditional'

for x). Therefore, no reduced MCMC runs were required (section 2.1 of Chib & Jeliazkov (2001)), only a single application of equation 5.12. However, the sampling step itself was inefficient relative to multiple block sampling because no Gibbs sampling was used so that the Metropolis step had a greater chance of being rejected.

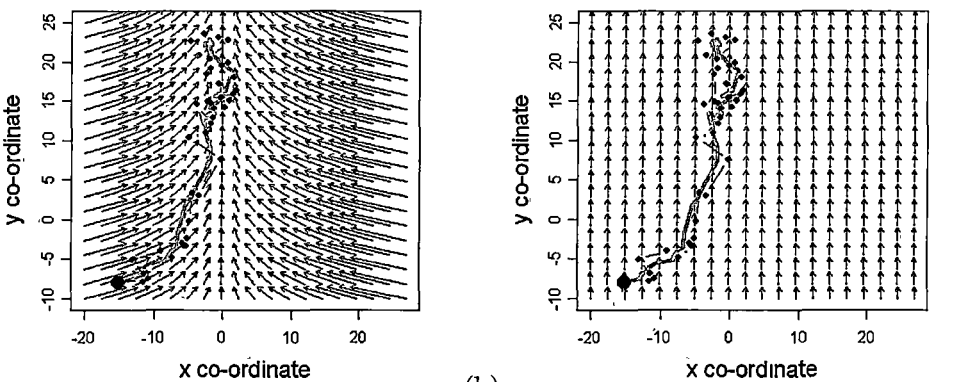
The first 1000 sets of draws were discarded from each MCMC run before performing an additional 100 000 iterations. The CODA diagnostic software and examination of the trace indicated that this was sufficient to reduce the influence of the starting point and to achieve convergence. Assessment of convergence included running several chains, each with different starting points and ensuring that the results did not differ noticeably among chains.

5.3.7 Spread of model comparison measures

The MCMC and model comparison calculations were repeated 5 times, providing 5 estimates of the marginal likelihood and the DIC for each scenario (in all cases the same 'observed' track was used) in order to calculate the coefficient of variation (c.v.) of each measure. Carlin & Louis (2000) terms this the 'brute force' method of assessing variance, but it is the only method available for DIC. The computationally intensive marginal likelihood estimates, and the DIC, showed little variation so that 5 repeats, while a small number, seemed adequate.

5.3.8 Simulation and estimation

The movement of a white shark moving northwards along the east coast of Australia, is simulated. It has been proposed that white sharks sometimes travel along lines of bathymetry (Barry Bruce, CSIRO, pers comm.). Therefore, a single track is simulated using an advection field that flows northwards but that has a lateral component towards a preferred N-S line of bathymetry (Figure 5.1). The support given by the simulated data to each of two models is compared, one using the advection field used to simulate the data, and the other using a constant northerly flow. The parameter values used for the simulation are shown in Table 5.2. The velocities v, which add inertia to the path, are taken to be relatively precise, with precision parameter $\tau^v = 0.01$. The path and the observations are allowed greater variability with precisions $\tau^x = \tau^y = 1$.



(a) (b)

Figure 5.1: Advection fields for (a) the simulation and (b) the incorrect alternative hypothesis. The simulated path x is shown as thick grey line and the starting point for the path x_1 as a large black dot. The simulated observations y are shown as small black dots.

Table 5.2: Parameter values and dimensions used in simulating the track.

Symbol	Description	Value
N	Number of locations	41
n	Number of observed locations	21
m	Number of advection fields	1
Λ	Step durations (drawn randomly)	$\sim$ N(0.75, 0.3)
$\mathcal{X}_i$	Advection forces	preferred depth constant northwards
β	Coefficient of advection	1.0
Px	Unobserved locations	$x_t, t = 2, 4, \cdots, 40$
x_1	Initial location	$(-15, -8)$
v_1	Initial velocity	$(0, 0)$
τ^y	Precision for observed locations y	$(1^{-1}, 1^{-1})$
τ^x	Precision for locations x	$(1^{-1}, 1^{-1})$
τ^v	Precision for velocity v	$(0.01^{-1}, 0.01^{-1})$

Three estimation scenarios were examined (Table 5.3). First, the gamma priors for the precision parameters for the path τ^x and the velocities τ^v have the correct mean and a somewhat diffuse prior whose variance is 1. Second, the priors have a mean that is 5 times the correct value and again a prior whose variance is 1. Third, the prior means are correct but the priors are very diffuse, having a variance of 100. In all scenarios, the prior for the precision for velocity is taken to have a mean of 0.01 and a variance of 1, because this parameter is considered to be a model setting rather than a parameter of interest. Note that the shape a and rate b parameters of the gamma distribution can be calculated from the distribution's mean μ and variance σ^2 using the formulae $a = \mu^2/\sigma^2$ and $b = \mu/\sigma^2$. Table 5.3 expresses the gamma prior distributions both in terms of shape a and rate b, and also mean and variance.

Table 5.3: Priors used for estimation.

Symbol	Description	Value
β_0	Mean of multivariate normal prior for β	0
τ_0	Precision for prior for β	1
x^0	Prior mean initial location	$(-15, -8)$
τ^{x0}	Precision for initial location	$(1000, 1000)$
v^0	Prio mean initial velocity	$(0, 0)$
τ^{v0}	Precision for initial velocity	$(1000, 1000)$
a^v	Scale for gamma prior for τ^v	$(1e^4, 1e^4)$
b^v	Rate for gamma prior for τ^v	$(100, 100)$
	Mean for gamma prior for τ^v	$(0.01^{-1}, 0.01^{-1})$
	Variance for gamma prior for τ^v	$(1, 1)$
Estimation scenario 1, for τ^x and τ^y		
$a^x = a^y$	Scale for gamma priors	$(1, 1)$
$b^x = b^y$	Rate for gamma priors	$(1, 1)$
	Mean for gamma priors	$(1^{-1}, 1^{-1})$
	Variance for gamma priors	$(1, 1)$
Estimation scenario 2, for τ^x and τ^y		
$a^x = a^y$	Scale for gamma priors	$(25, 25)$
$b^x = b^y$	Rate for gamma priors	$(5, 5)$
	Mean for gamma priors	$(0.2^{-1}, 0.2^{-1})$
	Variance for gamma priors	$(1, 1)$
Estimation scenario 3, for τ^x and τ^y		
$a^x = a^y$	Scale for gamma priors	$(0.01, 0.01)$
$b^x = b^y$	Rate for gamma priors	$(0.01, 0.01)$
	Mean for gamma priors	$(1^{-1}, 1^{-1})$
	Variance for gamma priors	$(100, 100)$

The prior for the coefficients β has a mean of zero, to reflect the prior assumption

that any advection field included in the model is not effecting movement, and has a precision of 1. The prior for the first location is centred on the observation of the first location and is taken to be very precise. The prior for the first velocity is similarly precise and has a mean at the origin $(0, 0)$ (Table 5.3).

5.4 Results

As the lengthy Methods section of this chapter attests, the chance of an implementation mistake in either coding or methodology of the variants on the Chib method, is high. The harmonic mean and Monte Carlo methods, although impractical for use with the full problem considered here, are much more easily implemented. The Monte Carlo method has the additional advantage of not making use of MCMC outputs, and therefore provides a wholly independent alternative estimate to that provided by the other methods. Therefore, in order to test the R code used in the full version of the model, a version was run that used precise priors centred on the true parameter values. The estimated marginal likelihood for all versions of the Chib method, the harmonic mean, and the Monte Carlo method all agreed to one decimal place.

When using more realistic, diffuse, priors the Monte Carlo method failed to converge even for chains of 1 000 000 iterations (Figure 5.2). The harmonic mean method appears to converge and give stable results across chains but does not return a correct estimate (Table 5.4). This failure has also been shown in Chapter 6 and by Rosenkrantz & Raftery (1994).

The Bayes factors computed using the Chib & Jeliazkov (2001) method for estimating the marginal likelihood provide positive to very strong evidence in support of the correct model over the model that ignores depth preference (Table 5.4) in the two scenarios that used somewhat diffuse priors (variances of 1) for the precision parameters for the path and the observations. It did so even when the mean of those priors was five times the correct value. However, when highly diffuse priors are used (variances of 100) the incorrect model is favoured (Table 5.4). Interestingly, the two Chib & Jeliazkov (2001) estimates of the marginal likelihood agree more closely and give stronger evidence in favour of the correct model in the scenario that has the priors for the precisions for the path τ^x and the observations τ^y centred on an incorrect value (scenario 2, Table 5.4). The scenario that uses highly diffuse priors for the precision parameters shows an even higher c.v., possibly indicating lack of

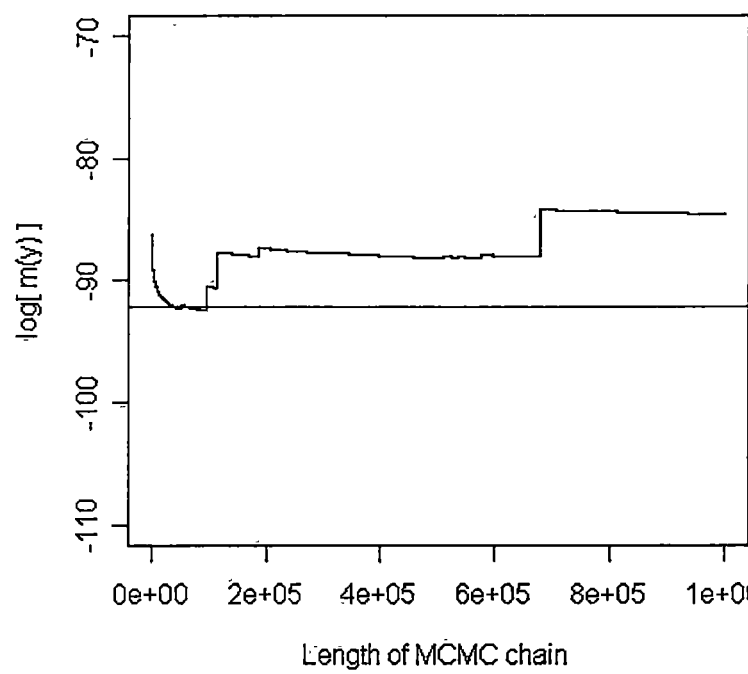


Figure 5.2: Trace of marginal likelihood estimated by Monte Carlo integration plotted against the number of MCMC draws made from the joint prior distribution. The estimate of the marginal likelihood made using Chib's method is shown, as a horizontal line, for comparison.

convergence, although this is not as great a problem over the parameter space of the alternative (incorrect) model (Table 5.4).

Table 5.4: Mean (and coefficient of variation c.v.) estimated marginal likelihoods and the Bayes factor for two competing models, one using the correct and the other an incorrect advection field. Estimation was by the multiple block (CJ), and single block (CJ1) Chib & Jeliazkov (2001) methods and the harmonic mean method (HM). The final column gives an evaluation of the evidence in favour of the correct model. Results are shown for the three estimation scenarios detailed in Table 5.3, which use different priors for the precision parameters τ^x and τ^y.

Estimator	log(Marginal likelihood)		log(Bayes factor)	Evidence for
	Correct	Incorrect		correct model
Estimation scenario 1: correct means, variances = 1				
CJ	-92.8 (2.6%)	-95.3 (0.3%)	2.5	Positive
CJ1	-90.5 (2.2%)	-95.1 (0.2%)	4.6	Strong
HM	-76.1 (0.9%)	-75.9 (0.9%)	-0.2	None
Estimation scenario 2: incorrect means, variances = 1				
CJ	-104.9 (1.2%)	-120.3 (0.03%)	15.4	Very strong
CJ1	-105.8 (1.2%)	-120.4 (0.1%)	14.6	Very strong
HM	-75.2 (1.2%)	-65.1 (3.2%)	-10.1	None
Estimation scenario 3: correct means, variances = 100				
CJ	-155.8 (17.0%)	-99.5 (0.1%)	-56.3	None
CJ1	-154.5 (12.2%)	-99.6 (0.4%)	-54.9	None
HM	-76.6 (1.5%)	-75.0 (5.7%)	-1.7	None

The Bayes factor computed using the harmonic mean estimator, favours the incorrect model in all scenarios, and shows greater c.v. than the other estimators. Chapter 6 shows that this estimator performs unacceptably in cases such as this where the prior is diffuse.

DIC identified the correct model in all three scenarios. It was relatively stable, having lower c.v., on the whole, than those for the marginal likelihood calculations (Tables 5.4 and 5.5).

Table 5.5: Deviance Information Criterion (DIC) for the three estimation scenarios detailed in Table 5.3.

Scenario	DIC		Difference
	Correct	Incorrect	
1	299.5 (0.2%)	349.5 (0.1%)	50.0
2	233.8 (0.1%)	296.4 (0.1%)	62.6
3	267.6 (1.9%)	329.2 (0.1%)	61.6

5.5 Discussion

The Bayes factor is used to compare a null model with any number of alternative models, and although it can also be used for Bayesian model averaging (Hoeting *et al.* , 1999), it differs in this way from AIC, BIC, and DIC, which compare a suite of models simultaneously without the need to identify one as the null model. It is given by the ratio of the marginal likelihoods of the two models. The marginal likelihood is the integral of the likelihood with respect to the prior across all possible values for the parameters. If the alternative (non-null) model has a marginal likelihood that is at least 3 times that of the null model on a log scale then it is considered that the data provide sufficient evidence in favour of this model over the null model (Kass & Raftery, 1995). A more diffuse prior, indicating greater ignorance, will give greater weight to parameter values with low corresponding likelihood so that the more diffuse the prior, the smaller the marginal likelihood. Bartlett's paradox (Bartlett, 1957) notes that such non-informative priors for the alternative models will cause the Bayes factor to favour the null model over the alternative (Kadane & Lazar, 2004). For this reason, it has been argued that diffuse priors, while appropriate for estimation problems, should never be used in conjunction with the Bayes factor (Aitkin, 1991). The closely related Lindley's paradox (Lindley, 1957) notes that if the prior is very diffuse, as the number of model parameters increases, less weight is placed on any alternative so that the marginal likelihood decreases such that the Bayes factor will favour more parsimonious models (Gelfand, 1996).

Model comparison using Bayes factor in conjunction with the modelling framework presented here, correctly identified the advection field (and therefore the hypothesis regarding movement behaviour) that was used to simulate the track data, except when a highly diffuse prior (variance=100) was used for the precision parameters for the observations and the path. The advice against using highly diffuse priors when using the Bayes factor is supported by these results. However, highly diffuse priors reflect uncertainty in model parameters. While studies such as that of Vincent *et al.* (2002) may provide an indication of the precision of the observations, allowing the use of an informative prior, it is unlikely that prior information will exist for the precision of the path, so that a diffuse prior is likely to be desirable. Movement behaviour is, as yet, poorly understood so that precise priors are unlikely to be appropriate for most parameters of models of movement behaviour.

The much less computationally intensive deviance information criterion DIC also identified the correct model in all scenarios, even when highly diffuse priors were used. It would seem to be preferable to use DIC for model comparison when priors

are, of necessity, diffuse. DIC is less sensitive to the choice of prior, is much easier to compute, and showed, in most cases, a low c.v. relative to the Bayes factors computed here. Unlike alternatives such as AIC and BIC, DIC does not require that the sample size or the number of parameters be known, which makes it a useful tool for application to hierarchical or mixed effects models, such as the model presented in this thesis. It is easily calculated from a set of MCMC draws and is not computationally taxing with regard to repeated integration of the likelihood.

Kass & Raftery (1995) recommend (prior to the appearance of DIC in the literature) that BIC, which requires less computation than the Bayes factor, be used as a rough approximation, and that calculation of the Bayes factor be reserved for more detailed analyses. Because of the sensitivity of the Bayes factor to the choice of prior, they recommend that, if the Bayes factor is used, sensitivity analysis should be performed using a reference set of priors.

Other forms of the Bayes factor are available that attempt to overcome some of their problems (in particular, sensitivity to the prior), although these typically have problems of their own (Kass & Raftery, 1995; Kadane & Lazar, 2004).

The work discussed in this thesis emphasizes model selection over parameter estimation, indeed the interest is not always in the parameter values themselves but rather in sifting less likely from more likely hypotheses. Prior information is likely to be lacking so that it seems preferable to use the less sensitive DIC in future work. However, DIC is sensitive to the parameterisation of the hierarchical model (Spiegelhalter *et al.* , 2002; Carlin & Louis, 2000) and, as discussed in Section 3.5, the modelling framework presented here allows for several alternative formulations, the choice of which is sometimes arbitrary in terms of biological realism. These may not prove to be arbitrary in terms of DIC, a potential problem that requires further investigation.

In practice, it is not expected that any hypotheses regarding movement behaviour and their associated advection fields will be correct, in terms of exactly representing the animal's true complex behaviour, but it is expected that the tools presented here will be successful in discriminating more likely hypotheses from less likely ones. This work is not concerned with exact and unbiased calculation of parameters or advection fields, but it is instead hoped to discern major trends in behaviour.

Chapter 6

The Harmonic Mean Estimator for the Marginal Likelihood is not Appropriate for Practical Model Comparison

This chapter is based on a manuscript that has been accepted by the Australian and New Zealand Journal of Statistics for publication as a short note.

6.1 Abstract

To calculate the Bayes factor, a Bayesian mechanism for model comparison, the marginal likelihood must be estimated. Of the numerical methods available for estimating this integral the most attractively simple, and widely used, is the harmonic mean method. Although its instability is acknowledged in the literature, it is said to be accurate enough for practical model comparison. A simple example is used to show that when the posterior is sufficiently lighter-tailed than the prior, this method fails, having insufficient accuracy to allow model choice using the Bayes factor. Furthermore, results can be highly inaccurate while giving the appearance of convergence.

Keywords: Bayes Factor; importance sampling; Markov chain Monte Carlo

6.2 Introduction: Bayes factor and the marginal likelihood

Model comparison is an important aspect of mathematical inference. In the case of Bayesian computation, the primary means of model comparison is the Bayes factor Jeffreys (1935); Kass & Raftery (1995). Its calculation involves a difficult integral - the marginal likelihood or normalising factor for the posterior. Only in very simple cases is it possible to calculate this quantity analytically, it must usually be estimated.

Although MCMC provides accurate methods for estimating the shape of the posterior, the normalising factor remains unknown. However, the set of draws from the posterior that MCMC does provide can be used to estimate the marginal likelihood. Many estimators have been proposed. These include the Laplace approximation (Tierney & Kaldane, 1986; Lewis & Raftery, 1997) and its variants (see Gilks *et al.* , 1996; DiCiccio *et al.* , 1997), Candidate's method (Chib, 1995; Chib & Jeliazkov, 2001) and four nested importance sampling estimators, bridge sampling (Gelman & Meng, 1998), path sampling (Meng & Wong, 1996), Reciprocal Importance Sampling (Gelfand & Dey, 1994) and the harmonic mean method (Newton & Raftery, 1994). Reviews can be found in DiCiccio *et al.* (1997) and Kass & Raftery (1995). The easiest of these to implement, the harmonic mean method proposed by Newton & Raftery (1994), requires nothing more than a set of independent draws from the posterior.

Although the inherent instability of the harmonic mean estimator is acknowledged in the literature, it is said to be sufficiently accurate for practical model comparison (notably Rosenkrantz & Raftery (1994); Carlin & Chib (1995); Kass & Raftery (1995); Raftery (1996); Sorensen & Gianola (2002)). This has lead other authors to use the method in good faith e.g. Rivot *et al.* (2004) write *'It is well known that the (harmonic mean estimator)...may be quite unstable...However, this approximation is likely to give results that are accurate enough for the interpretation in the logarithmic scale (Kass and Raftery, 1995)'*. They proceed to use this estimator and no other to perform model comparison. The harmonic mean is also the only method of model comparison included in a widely used statistical package (MrBayes: Huelsenbeck & Ronquist (2001); Ronquist & Huelsenbeck (2003)).

The suggestion that the harmonic mean estimator is adequate for practical model comparison is disputed. An example is presented that shows that this estimator

is inadequate even in a trivial case. Moreover, the method is weakest when the data are most informative. Little confidence should be placed in this estimator. Despite repeated warnings regarding instability this point is not made in the literature.

The harmonic mean method is contrasted with the next most complex importance sampling estimator, Reciprocal Importance Sampling RIS (Gelfand & Dey, 1994) and conclude that if simplicity is required RIS is to be preferred.

6.3 Methods

6.3.1 Importance sampling the marginal likelihood

Both the harmonic mean and RIS estimators can be derived by noting that for any density $f(\theta)$

$$1 = \int f(\theta)\, d\theta = \int \frac{f(\theta)}{p(\theta \mid y)} p(\theta \mid y)\, d\theta. \tag{6.1}$$

Applying Bayes Theorem to the denominator yields

$$1 = \int \frac{f(\theta)\, m(y)}{p(y \mid \theta)\, p(\theta)} p(\theta \mid y)\, d\theta$$

so that

$$m(y) = [\int \frac{f(\theta)}{p(y \mid \theta)\, p(\theta)} p(\theta \mid y)\, d\theta]^{-1}.$$

If $\theta^{(i)}$, $i = 1, \ldots, n$ are independent draws from the posterior $p(\theta \mid y)$, the Monte Carlo estimate is

$$\hat{m(y)} = \frac{1}{N} \sum_{i=1}^{N} \left[\frac{f(\theta^{(i)})}{p(y \mid \theta^{(i)})\, p(\theta^{(i)})} \right]^{-1}. \tag{6.2}$$

Although the aim is to solve $\int f(\theta)\, d\theta$, MCMC provides draws from the posterior which is therefore, in this case, the importance sampling density. The success of importance sampling rests on the congruence of the importance sampling density and the density distribution of θ, in this case $f(\theta)$. Efficiency derives from their similarity; accuracy from a guarantee that $p(\theta \mid y)$ is more diffuse than $f(\theta)$ across

the domain of θ (Geweke, 1989). For a Central Limit Theorem to apply to the estimator the importance sampling density, in this case $p(\theta \mid y)$, must be more diffuse than $f(\theta)$ (Geweke, 1989).

The harmonic mean estimator results from choosing $f(\theta)$ to be the prior $p(\theta)$

$$\hat{m(y)} = \left[\frac{1}{N}\sum_{i=1}^{N} p(y \mid \theta^{(i)})^{-1}\right]^{-1}.$$

Clearly a poor choice. If the data contains any information at all, the prior will be more diffuse than the posterior. Consequently Kass & Raftery (1995) have described the harmonic mean estimator as lacking a central limit theorem.

The RIS estimator, equation 6.2, leaves the choice of $f(\theta)$ to the investigator. The most efficient choice would be the posterior itself however it is important to err on the side of caution in ensuring that $f(\theta)$ is less diffuse than the posterior for all θ.

6.3.2 Simulation

To evaluate the harmonic mean and RIS estimators, consider the simple example of i.i.d. data y_i, $i = 1, \ldots, n$,

$$y_i \sim \mathrm{N}(\mu, 1)$$

with normal prior

$$p(\mu) \sim \mathrm{N}(\mu_0, \tau_0)$$

the posterior is

$$\mu \mid y \sim \mathrm{N}(\hat{\mu}, \hat{\tau}) \tag{6.3}$$

where

$$\hat{\mu} = \frac{\mu_0 \tau_0 + \sum_{i=1}^{n} y_i}{n + \tau_0}$$

and

$$\hat{\tau} = \tau_0 + n.$$

Direct integration yields the marginal likelihood

$$m(y) = (2\pi)^{-n/2} \sqrt{\frac{\tau_0}{\tau_0 + n}} \exp\left[-\frac{1}{2}\left(\sum y_i^2 + \tau_0 \mu_0^2 - \frac{(\tau_0 \mu_0 + \sum y_i)^2}{\tau_0 + n} \right) \right].$$

Values of $\mu = 0$, $\mu_0 = 0$, and $\tau_0 = 0.1$ were used and the simulation was repeated for four data sample sizes $n = 1$, 5, 50, and 500. The RIS estimator was applied using, for $f(\mu)$ (equation 6.1)

$$\mu \sim N(\hat{\mu}, 1.5\,\hat{\tau})$$

($\hat{\mu}$ and $\hat{\tau}$ given by equation 6.3). The most efficient choice for $f(\mu)$ is the posterior (Meng & Wong, 1996) but the more realistic situation is simulated in which the posterior is unknown and the operator errs on the side of caution by using a thinner-tailed distribution.

R code (R Development Core Team, 2007) for this simulation is given in Appendix 15.

6.4 Simulation results

Figure 6.1 shows a single MCMC chain for each of four data sample sizes. Percentiles of the results from 100 chains are shown in Table 6.4. The harmonic mean operator

performs poorly, and its performance worsens as the sample size of the data increases (Figure 6.1). Large 'jumps' occur whenever a parameter value is drawn from the tails of the posterior distribution and there is little sign of convergence to the true marginal likelihood value even after 1000 000 posterior draws. By contrast, the RIS estimator converges quickly and smoothly so that it is difficult to distinguish graphically.

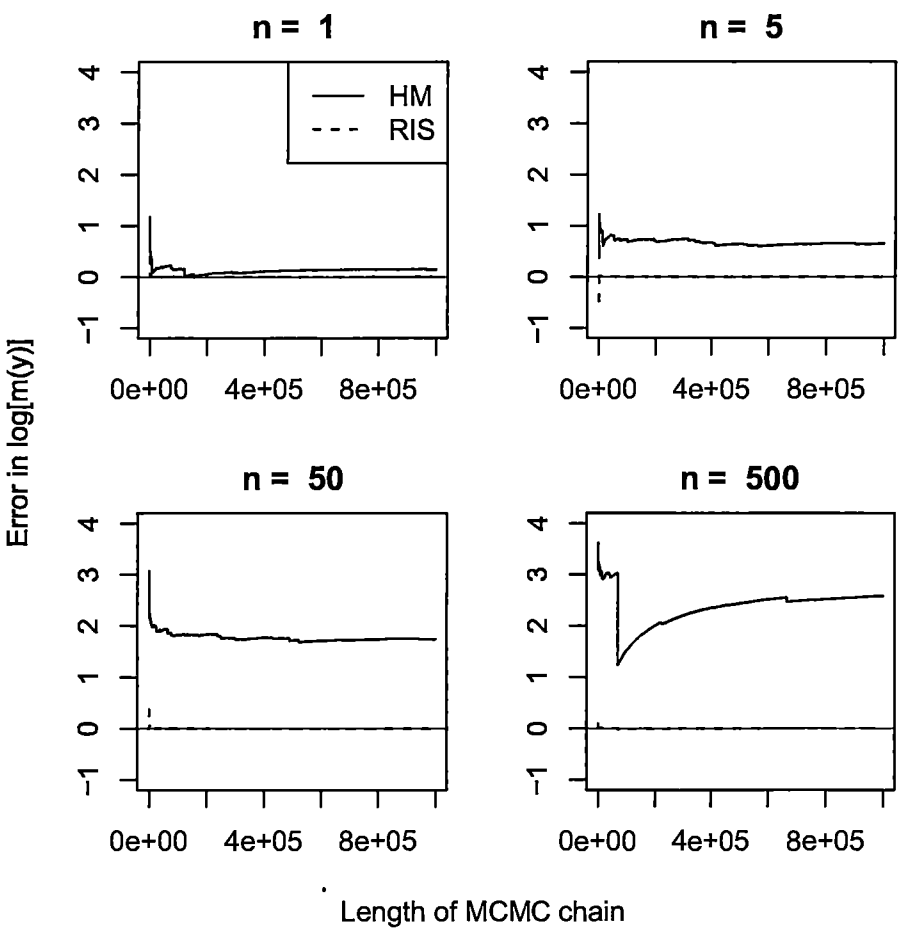


Figure 6.1: Error in the estimated log marginal likelihood for an increasing number of draws from the posterior distribution is shown. Results are shown for four data sample sizes n. The harmonic mean (HM), and the Reciprocal Importance Sampling (RIS) estimators are shown.

Table 6.1: Showing 5, 50 and 95 percentiles from 100 repeated estimations of the error in the marginal likelihood $(\hat{m(y)} - m(y))$. Each estimation $\hat{m(y)}$ used 1000 000 draws from the posterior. Results are shown for four values of n for the harmonic mean and the RIS estimators.

n	Harmonic mean			RIS		
	5	50	95	5	50	95
1	-0.07	0.13	0.21	-0.003	0.0004	0.003
5	0.24	0.62	0.73	-0.003	0.0003	0.003
50	1.01	1.70	1.79	-0.004	0.0001	0.003
500	2.50	2.84	2.94	-0.003	0.0005	0.003

6.5 Discussion

The harmonic mean estimator fails because it violates a central requirement of importance sampling: that the importance sampling density have heavier tails than the density distribution of θ (Geweke, 1989). The posterior distribution must have lighter tails than the prior if the data contain any information at all and is therefore *always* a poor choice for the importance sampling density. Yet this is the distribution used by the harmonic mean estimator. Consequently Kass & Raftery (1995) state that the harmonic mean estimator lacks a Central Limit Theorem, in other words, estimation precision does not increase with number of draws made. A value from the extremes of the likelihood distribution may be drawn at any time resulting in a large inverse value and consequently a large 'jump' in the harmonic mean estimate (Figure 6.1). The estimator is at its worst when the data are more informative (in this case a larger sample size) or the prior is more diffuse. The harmonic mean estimator will perform best when the prior and posterior are most similar. Hsiao *et al.* (2000) used priors based on actual prior information and found that the harmonic mean gave similar results to those of other estimators. By contrast in a comprehensive simulation Lopes & West (2004) found that the harmonic mean was one of few estimators that performed poorly. Poor performance has been shown here.

A number of authors (Rosenkrantz & Raftery, 1994; Carlin & Chib, 1995; Raftery, 1996) state that the harmonic mean estimator is sufficiently accurate for evaluating the Bayes factor on the scale presented by Jeffreys (1961), see Table 2. Here a difference of more than 1 in the marginal likelihoods is considered to be positive evidence. In this simple example the harmonic mean estimator produced errors of this magnitude in most simulations (Table 1). Furthermore, by equation 5.1 the error in the marginal likelihood may be doubled when computing the log Bayes factor.

The insidious aspect of this estimator is the apparent stability of the biased result shown in Figure 6.1. This is very likely to be taken as an indication that the harmonic mean estimate has converged to the correct value.

Many good, alternative estimators exist (see DiCiccio *et al.* (1997); Sorensen & Gianola (2002)) including very good Monte Carlo estimators (Meng & Wong, 1996; Gelman & Meng, 1998) however the harmonic mean estimator survives, despite its known instability, because of its simplicity.

The RIS method performs well provided the importance sampling density has lighter tails than the posterior. In practical implementation it differs from the harmonic mean method only in the need to select the importance sampling distribution. This might require an additional MCMC calculation in order to estimate an approximation to the posterior. However, because the importance sampling density need not be a close approximation to the posterior, just lighter-tailed, crude approximation from a short MCMC run can be offset by a greater reduction in the heaviness of the tails. The need to make an appropriate choice is a limitation of the method (Chib, 1995; DiCiccio *et al.* , 1997) but a poor choice only causes the RIS estimator to suffer the failing already inherent in the harmonic mean estimator.

6.6 Conclusions

The harmonic mean estimator will always be unreliable because the posterior distribution will always be less diffuse than the prior. Its performance will worsen, however, as the ratio of the prior to the posterior variance increases. This can arise through greater information content in the data or through the use of diffuse priors. Furthermore, apparent stability in the result is no guarantee of convergence.

This Chapter proposes one alternative to the harmonic mean estimator however many others exist (see DiCiccio *et al.* (1997) for a thorough review). The RIS estimator, while performing well in simulation tests (Newton & Raftery, 1994; Kass & Raftery, 1995) does require a good approximation of the posterior and therefore might not always be the preferred estimator. However, it will always be preferable to the harmonic mean estimator.

Previous simulation tests may have been conducted with optimistically informative

priors and greater caution is suggested when recommending the harmonic mean estimator. The RIS estimator is better and hardly more complicated. It has both a Central Limit theorem and finite variance (Chib, 1995). The support for practical application of the harmonic mean estimator in the literature seems unjustified.

Chapter 7

Using Laplace's equation to guide moving individuals around obstacles.

7.1 Abstract

A frequently encountered problem when modelling the movements of animals occurs when modelled animals become trapped in semi-enclosed parts of their landscape, or cross impassable barriers that are narrower than their step length. The navigational methods used by real animals to prevent entrapment in semi-enclosed areas such as bays (for marine animals) are not well known. Such navigational methods may be complex and if they are not the focus of the modelling study then their inclusion in a model would not be desirable. Few published studies discuss the methods used to overcome the problem of entrapment. Those that do, allow animals to die when they meet a barrier, or to rebound for several steps.

Models applied to satellite tracking data typically involve few individuals so that these cannot be allowed to die on encountering a barrier. The rebound rule has worked well in individual-based modelling studies, but it would be difficult to incorporate into a statistical model that requires analytical evaluation of the equations governing behaviour. This chapter presents an alternative solution, using an advection field to guide moving animals past barriers.

This advection field is derived using Laplace's equation for the diffusion of heat (or a chemical substance) through a medium. A target location is identified towards which the animal moves - this is treated as a heat source. Barriers (such as land, for a marine animal) are treated as insulators so that heat flows around these barriers. An advection field is derived by taking the first differences of the equilibrium heat

values. Two boundary conditions, Neumann and Dirichlet, are considered for the boundaries at the edge of the barriers and at the edge of the landscape. Dirichlet conditions (where the boundary takes on a specific value, zero heat in this case) applied to the barrier edge, cause animals to be directed away from the barrier whereas Neumann conditions for the barrier edge allow animals to skirt closely past the barrier. Which of these is more appropriate will depend on the nature of the animal under investigation. Some marine animals, such as white sharks, do enter shallow coastal waters whereas others, such as blue sharks, do not. It is not desirable to couple a Dirichlet barrier edge condition with a Neumann landscape edge because this causes simulated animals to be pushed towards the landscape boundary.

Simulations are presented for artificial landscapes, illustrating the importance of selecting a sufficiently large landscape, or fine grid scale and showing the effects of different choices for boundary conditions. A real example, using a section of the South Australian coast is used to contrast the results of this Laplace-advection method with not using any method, using a rebound rule in which the animal retraces its steps, or using a rebound rule in which the animal moves in a random direction.

Although the focus of this investigation was on presenting a solution to the problem of guiding animals past obstacles, the use of Neumann boundary conditions could present a partial solution to the problem of animals stepping across narrow barriers because it directs animals away from such barriers. This chapter presents an elegant method that can be used to guide animals past impassable barriers, towards a target location.

7.2 Introduction

Few investigators who have modelled the paths of individuals moving through a landscape have escaped the problem of their simulated individuals becoming trapped against obstacles in the landscape. Individuals that are moving blindly towards a target location will become stuck when they encounter an object in their path. Those individuals that use simple avoidance rules that take them around smooth obstacles can nevertheless become trapped in semi-enclosed parts of the landscape. Another problem occurs when the moving object steps over what was intended to be an impassable obstacle. This occurs when the width of the obstacle is smaller than the length of a step. Whilst these problems are mentioned informally, they seldom appear in the literature. Perhaps authors regard the problem as too trivial

to mention or perhaps it is the various solutions developed on the fly that are too ad hoc to draw attention to.

I have encountered these problems in the course of modelling the movement of marine animals through a landscape in which coastlines provide impassable barriers, bays present opportunities for entrapment and small islands and narrow peninsulas present themselves for stepping across. In the rest of this paper the term 'object' will be replaced with 'animal' and 'obstacle' with 'land' but this work is equally applicable to any particles actively moving through an environment containing barriers.

Real animals must have strategies for escaping entrapment. Ecological concepts such as the desire to maximize fitness and minimize risk will inform the animal's decisions. Its ability to remember past experiences and characteristics of its environment will inform these decisions. For some animal groups these escape strategies are reasonably well known. The movements of small terrestrial insects are relatively easy to study, consequently these are well understood. Ants have been found to use dead reckoning and memorized landmarks (Grocott, 2003), foraging butterflies conform to a simple random walk (Root & Kareiva, 1984). Jeanson *et al.* (2003) found that beetles placed in a container, followed a random movement pattern until they encountered the wall of the container. There they became trapped, following the wall endlessly. In most natural circumstances, wall following would presumably eventually lead to escape.

Models that simulate the movements of animals require rules that prevent unrealistic behaviour such as entrapment or barrier crossing. Wiktorsson *et al.* (2004) used a reflection condition so that the collembolan insects they were modelling rebounded when meeting an obstacle. Murphy *et al.* (2004), modelling the seasonal distributions of krill, allowed modelled krill to become trapped whenever they met the coast or ice-edge (a no-slip boundary in fluid dynamic terms). Because they were modelling large numbers of individuals, enough always escaped so that a meaningful distribution could be formed. Broekhuizen *et al.* (2003) guided their foraging albatross using a series of waypoints. Modelled birds fly to the nearest waypoint that will not require that they fly over land. Given sufficient waypoints on the map, this strategy allowed the birds to move to their observed foraging areas and to return home. Tracked birds were observed to make sudden large changes of direction between straight flight legs, so that the modelling strategy seemed to mimic that used by real birds. Bennett & Tang (2006) give a thorough review of ways in which agent-based models can represent the decision-making process and spatial knowl-

edge of a moving animal. However, none of these concepts deals with the minutiae of moving an animal past an obstacle in its path.

Birds and marine animals are more difficult to follow than terrestrial insects. For example, studies of birds typically discuss 'vanishing direction', the direction in which the birds were flying when they were last seen, rather than flight path. Recently, however, satellite-tracking devices have opened a window on the movements of animals large enough to carry them, allowing modellers the opportunity to investigate questions regarding their overall navigational strategies, many of which exist as unproved theories regarding how these animals might be navigating. Proposed navigational strategies include a magnetic compass (Walker *et al.* , 2002; Cain *et al.* , 2005; Lohmann & Lohmann, 2006), celestial compass (Gould, 1998), olfactory sense (Wallraff, 2004; Pascual & Quinn, 1991), and memorized landmarks (Benhamou *et al.* , 1995).

A model of navigation using memorized landmarks is unlikely to encounter problems of entrapment within the landscape. However, models that investigate whether animals are following a compass, such as a magnetic or celestially derived compass, probably would. The focus of such an investigation is likely to be whether the overall path taken by the animal shows anomalies that can be explained by one of these compasses and not another. For example, the earth's magnetic field has persistent irregularities; animals using a geomagnetic compass would not take the shortest distance between two points but would follow a theoretically predictable curve. Such deviations would allow an investigator to place greater weight on one hypothesized compass over another. The exact mechanism whereby such an animal extracted itself from a bay or moved around an obstacle would not be of interest. Clearly, animals must have some strategy for moving along coastlines until they have passed the obstructing land and are able to resume their preferred course. Pascual & Quinn (1991) showed, using observations and models of migrating salmon, that directed strategies such as an ability to distinguish waters from different rivers by scent are needed to explain observed navigational success.

Unless the actual strategy that the moving individual employs to avoid obstacles is of interest to the modeller, it is not necessary to exactly replicate that strategy in simulation. If the aim is to investigate a navigational strategy such as a form of compass derived by sensing external factors, the simulation need only accurately replicate the emergent behaviour of the animal with respect to obstacle avoidance. Thus a clumsy ad hoc solution, which might not be worthy of drawing attention to in the publication of the model, might be sufficient.

This paper proposes an elegant, general, solution to this problem that uses Laplace's heat (diffusion) equation to generate an advection field that flows towards a point of attraction while flowing smoothly around obstacles. A marine example is considered in which an animal moves towards a target location in the ocean, using the (Laplace) advection field to navigate around intervening landmasses.

7.3 Methods

The problem is considered of an animal navigating across a landscape (terrestrial or aquatic) towards a particular location, its goal. The goal is considered to be a heat source, so that a temperature gradient occurs across the landscape. Obstacles to movement that occur in the landscape are considered to be temperature insulators. A moving animal can navigate towards the goal by following the temperature gradient, moving towards greater warmth. Doing so will guide the animal around the obstacles.

A grid is imposed on the landscape, dividing it into cells, some of which the animal may enter (water) and others not (land). This method is applicable to any landscape, and grid resolution is chosen by the investigator to suit their particular problem. A target cell is identified and this constitutes the moving animal's goal.

The simulation presented in this chapter used the C++ programming language for rapid calculation of equilibrium temperature states, and the R programming environment (R Development Core Team, 2007) for subsequent simulations of moving individuals and for graphical outputs.

7.3.1 Heat diffusing through a grid

The partial differential equation governing the spread of heat h through a conducting 2-dimensional medium (with spatial co-ordinates x and y) is the same as that governing the diffusive spread of particles (or randomly walking individuals) (Okubo & Levin, 2001; Murray, 1989)

$$\frac{\partial h}{\partial t} = k \left(\frac{\partial^2 h}{\partial x^2} + \frac{\partial^2 h}{\partial y^2} \right). \tag{7.1}$$

In a diffusion example, h would represent the number of particles, or moving individuals, and the rate constant k would be termed the diffusion constant D.

At equilibrium $\frac{\partial h}{\partial t} = 0$, and the system is governed by Laplace's equation

$$\frac{\partial^2 h}{\partial x^2} + \frac{\partial^2 h}{\partial y^2} = 0. \tag{7.2}$$

Laplace's equation can easily be solved on a 2-dimensional grid by applying Jacobi's method to the finite differences. This effectively results in the temperature of each cell being the average of that of its four cardinal neighbours, and the problem can be solved iteratively. The cell containing the heat source is given a constant temperature value, and the temperature of every remaining cell is replaced by the average of its four cardinal neighbours; this process is repeated until the system reaches equilibrium.

7.3.2 Boundary conditions

At a boundary, one may either assume Dirichlet conditions (the boundary takes on a specified value, $h = 0$) or Neumann conditions (the flux at the boundary takes on a specified value, $\frac{\partial h}{\partial t} = 0$), (Mayers & Morton, 2005). It is not clear which is the more appropriate, therefore both are considered.

There are two types of boundary cell to consider: the cells on the outermost edge of the grid (the *grid boundary*), and the land cells that are adjacent to water cells (the *land boundary*).

In the Dirichlet case, all boundary cells take on a temperature of zero. In the Neumann case, the land boundary cells take on the average value of their non-land cardinal neighbours and the grid boundary cells, which have only one non-boundary cardinal neighbour, take on its value. This leads to four scenarios

(a) both boundaries are Dirichlet;
(b) a Dirichlet land boundary and a Neumann grid boundary;
(c) a Neumann land boundary and a Dirichlet grid boundary; and
(d) both boundaries are Neumann.

The final scenario (d) leads to an accumulation of heat within the landscape so that at equilibrium, all cells have the same temperature. Therefore there is no temperature gradient from which to calculate advection forces. This scenario is not considered further.

7.3.3 Advection from temperature gradient

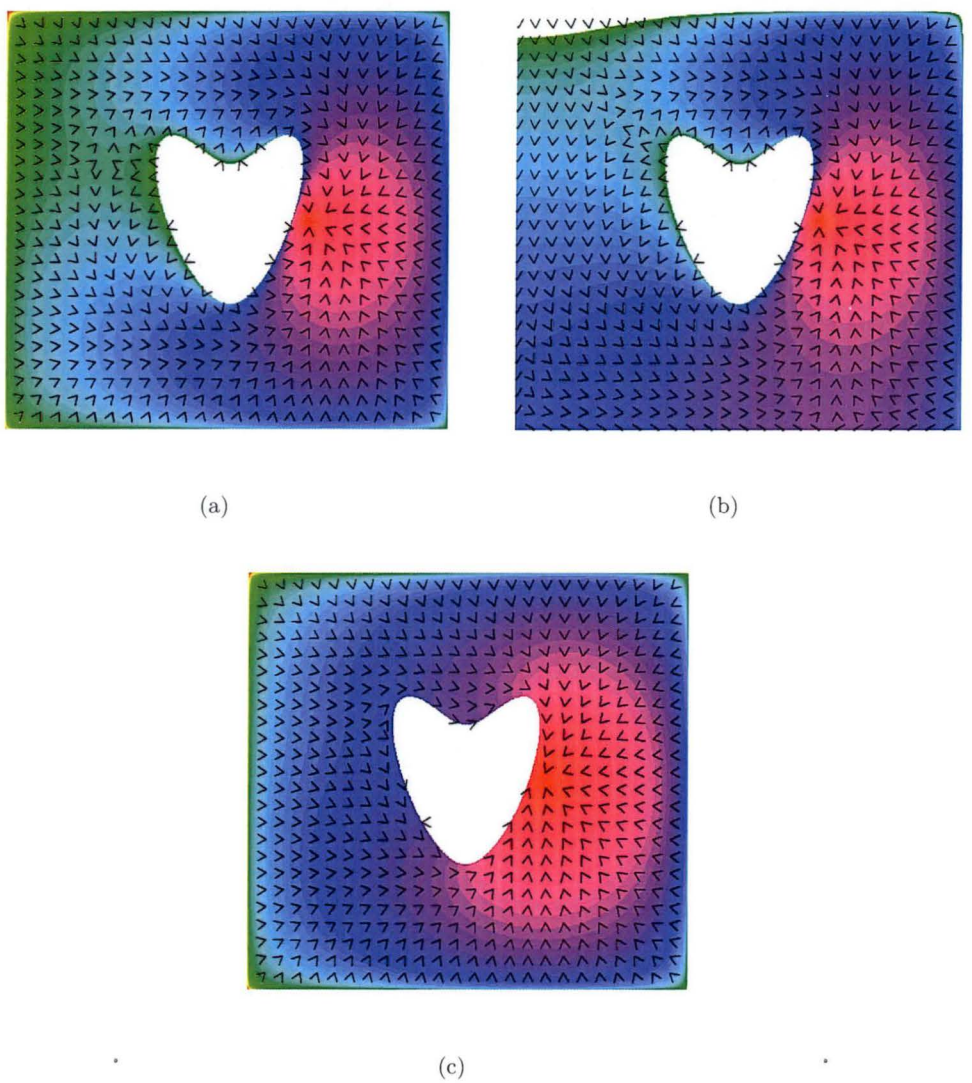
The equilibrium temperature grid was converted into vectors for a moving animal to follow, by differencing adjacent cells in each of the x- and y-directions. This yielded a set of vectors (U, V) corresponding to an advection field. As the aim is to use the temperature gradients to give direction, not speed, to an animal's movements, these vectors were converted to unit vectors (u, v) by dividing each by its length $\sqrt{U^2 + V^2}$

$$u = \frac{U}{\sqrt{U^2 + V^2}}, \quad v = \frac{V}{\sqrt{U^2 + V^2}}. \tag{7.3}$$

7.3.4 Simulated landscapes: central landmass

A landscape containing a central landmass surrounded by water was simulated on a 500 by 500 cell grid. One of these cells, to the northeast of the landmass, was allocated as a heat source - the goal towards which an animal would move. The method described in Section 7.3.1 was used to generate an advection field corresponding with each of the first three boundary condition scenarios listed in that section (Figure 7.1).

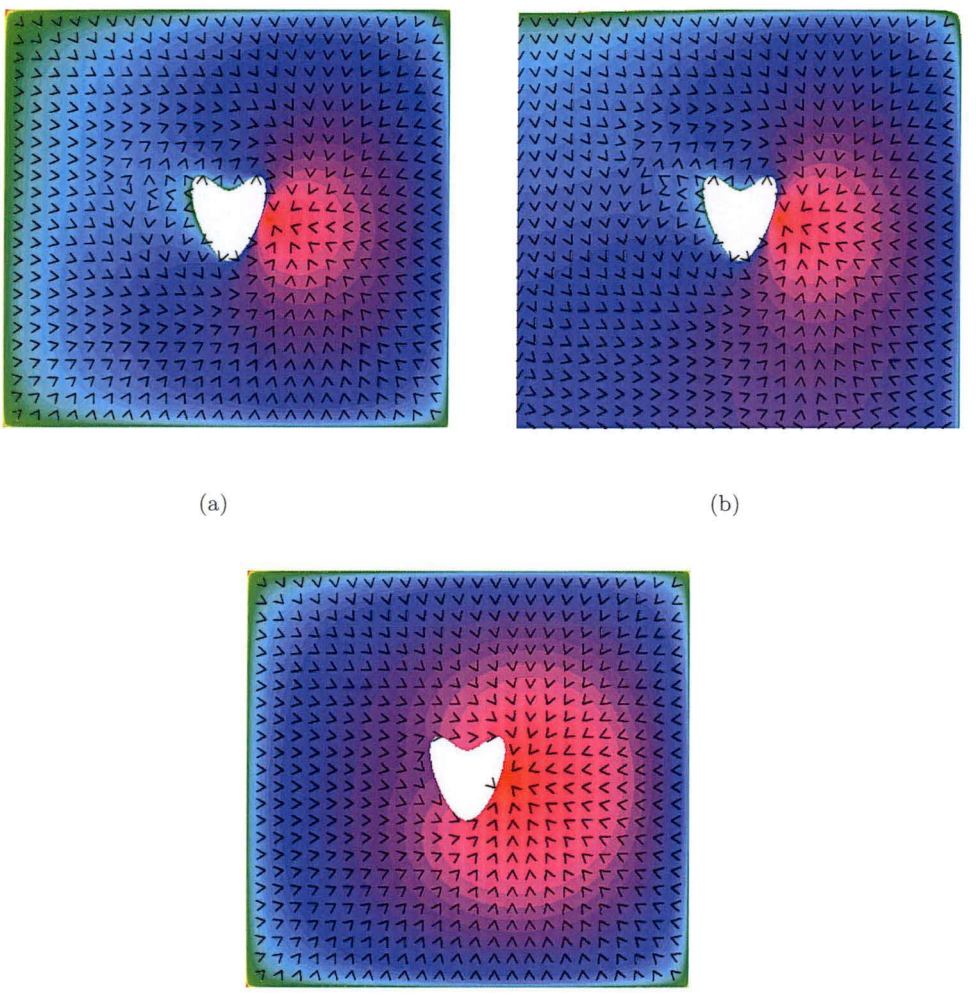
Another set of three advection fields was calculated, again corresponding with the four boundary conditions, but for a 500 by 500 grid containing a smaller landmass (Figure 7.2). This is equivalent to placing a coarser grid on a larger area, so that the landmass is less dominant on the grid.



(a) (b)

(c)

Figure 7.1: Advection fields generated from a temperature gradient around a simulated landmass that dominates the landscape. A 500 by 500 cell grid was used. Boundary conditions were (a) Dirichlet land and grid; (b) Dirichlet land and Neumann grid; and (c) Neumann land and Dirichlet grid. Colours indicate temperature on a log scale (hottest cells are red, coolest green).



(a) (b)

(c)

Figure 7.2: Advection fields calculated from a temperature gradient around a relatively small central landmass. A 500 by 500 cell grid was used. White indicates land, colours indicate temperature on a log scale (hottest cells are red, coolest green). Boundary conditions were (a) Dirichlet land and grid; (b) Dirichlet land and Neumann grid; and (c) Neumann land and Dirichlet grid.

7.3.5 Simulated landscape: South Australia

The problem of entrapment within semi-enclosed bays was considered, using a section of the South Australian coast to illustrate how the use of Laplace's equation can prevent simulated animals becoming trapped in bays. Bathymetry data collected by Geosciences Australia (Petkovic & Buchanan, 2002) was used to define the coastline of a section of South Australia where Australian white shark researchers commonly capture, tag, and release sharks for tracking purposes. The seal colonies on the Neptune Islands (Figure 7.3) are seen to be a focal point for white sharks. From here, tagged sharks swim either eastwards or westwards, apparently following the coastline or a line of bathymetry, and later return to the Neptune Islands (Bruce *et al.* , 2006). Those that return from the east must negotiate Coffin Bay (Figure 7.3) and although this a notorious spot for shark attacks (Bruce *pers comm*) it is not believed that real sharks become trapped in the bay in the way that those simulated using a simple movement model do.

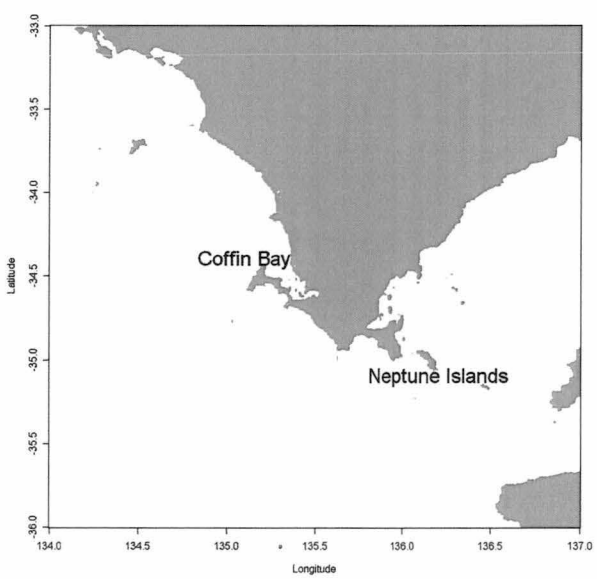


Figure 7.3: Part of the gulf region of South Australia, grey indicates land and white ocean.

In order to contrast movement using temperature-derived advection fields, with simple, naive navigation towards a target location, an advection field was derived that would move an animal from any point on the landscape directly towards a target cell placed in the centre of the Neptune archipelago (Figure 7.4). This is equivalent to an animal using an accurate compass sense to move directly towards a target

location.

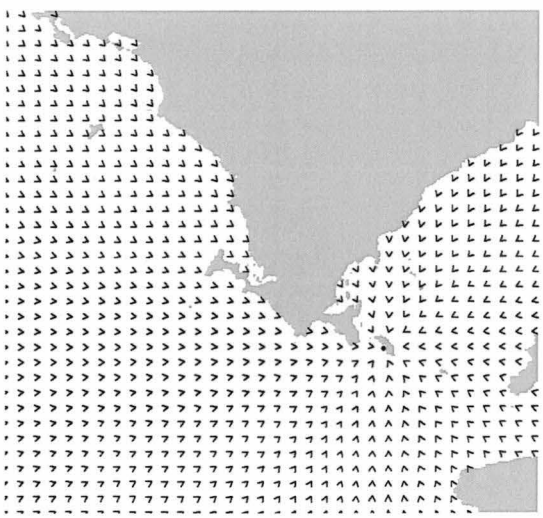


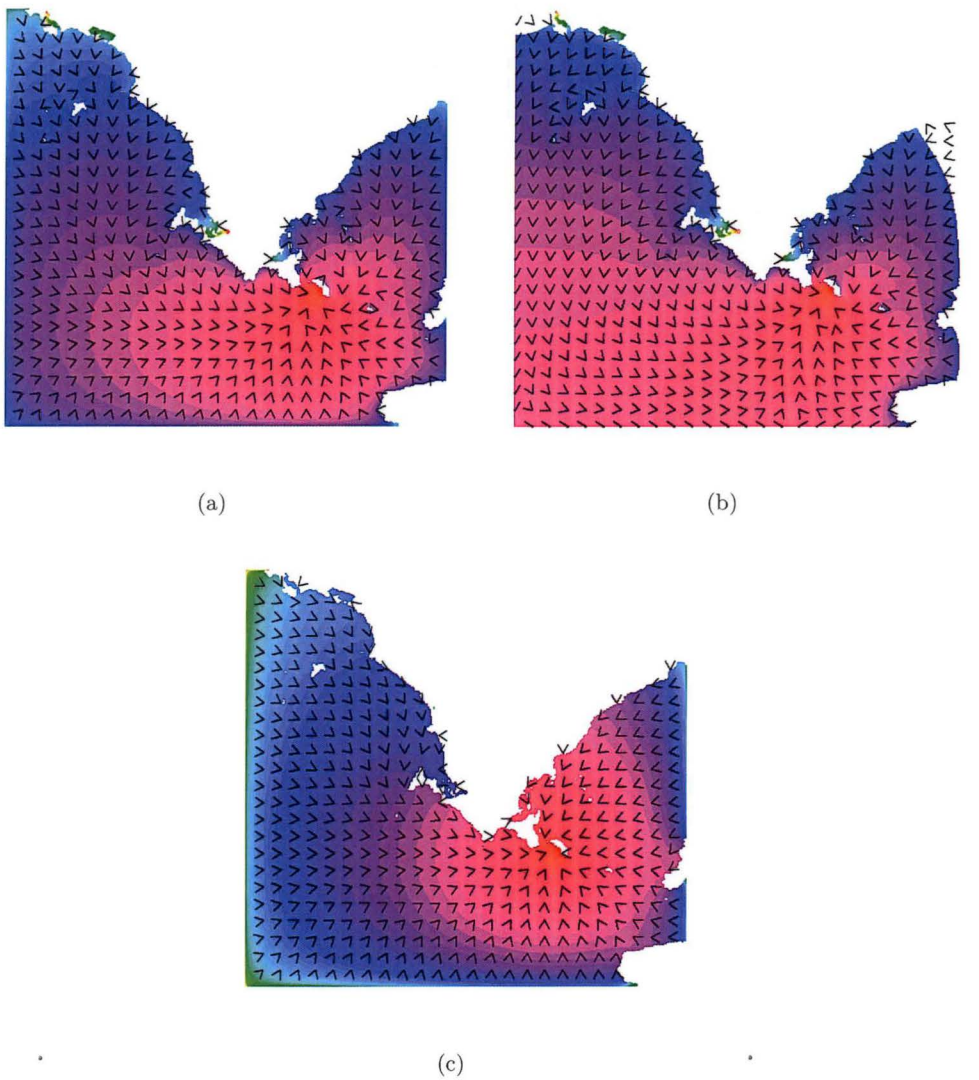
Figure 7.4: Advection directly towards the Neptune Islands

By contrast, the temperature-derived advection method was used to generate advection fields that flow, from the ocean, around the islands, towards the target cell in the centre of the Neptune archipelago (Figure 7.5). Again, a 500 by 500 cell grid was used.

7.3.6 Simulated movement

Simulations were performed of the movements of ten animals from a point in the ocean, westwards of the landmass shown in Figures 7.1 and 7.2, towards the goal (heat source) cell in the northeast, around the intervening landmass. Animals were started at a location in the northwest (with slight variation in their starting point). Movement proceeded by simple random motion, but with an additional advection component, so that the animal's location during time step k, $x_k = (x_{k1}, x_{k2})$ is a function of the previous location x_{k-1}, plus advection $X_{k-1} = (u_{k-1}, v_{k-1})$ at the previous location, and random error

$$x_k = x_{k-1} + X_{k-1} + \xi_{k-1}. \tag{7.4}$$



(a) (b)

(c)

Figure 7.5: Advection fields for part of South Australia calculated from a temperature gradient. A 500 by 500 cell grid is used. White indicates land, colours indicate temperature on a log scale (hottest cells are red, coolest green). Boundary conditions were (a) Dirichlet land and grid; (b) Dirichlet land and Neumann grid; and (c) Neumann land and Dirichlet grid.

Each component of the random error $\xi_k = (\xi_{k1}, \xi_{k2})$ has a normal distribution with variance σ^2

$$\xi_{ki} \sim N(0, \sigma^2), \quad i = 1, 2. \tag{7.5}$$

Each step taken by an animal was simulated by making a draw from this normal distribution but if the proposed location was on land, then it was rejected and another draw made, until a valid (non-land) location was obtained. The animal was placed at the centre of the cell into which this location fell. This introduces a degree of approximation to a true random walk so that the effects of the grid resolution need to be considered when interpreting the results of the simulation.

For most simulations, a variance of $\sigma^2 = 2^2$ was used, but the effect of altering the step size of the animals was examined by using variances of $\sigma^2 = 1^2$ and $\sigma^2 = 3^2$.

Finally, the effect of temperature-derived advected movement is contrasted with more conventional simulations in which animals move directly towards a target, using simple rules to avoid obstacles. In this case equation 7.4 was used again but the temperature-derived advection field was replaced by one that directed the animal towards the target location by the shortest route, ignoring the intervening land. First, animals walked towards the target with no additional rules should they encounter an obstacle. Second, a rebound rule was used - if the mean of the normal distribution described equations 7.4 and 7.5 was on land, then the animal moved in the reverse direction of advection for the next 30 steps. Third, once an animal encountered land it chose a random direction for each of the next 30 steps. After 30 steps, the animals resumed their original course towards the target cell. The number 30 was chosen as a compromise that gave the animal a chance of escape without unrealistically long digressions.

7.4 Results

7.4.1 Central landmass

Boundary conditions have a profound effect on the advection flow (Figure 7.1). When the land has a Dirichlet boundary (Figure 7.1a and b) the animal is repelled from the land and an area of divergence forms to the northwest of the landmass.

When the land boundary is Neumann (Figure 7.1c) advection in this region is more laminar. When the grid boundary is Dirichlet (Figures 7.1a and c) the animals are repelled from the grid boundary.

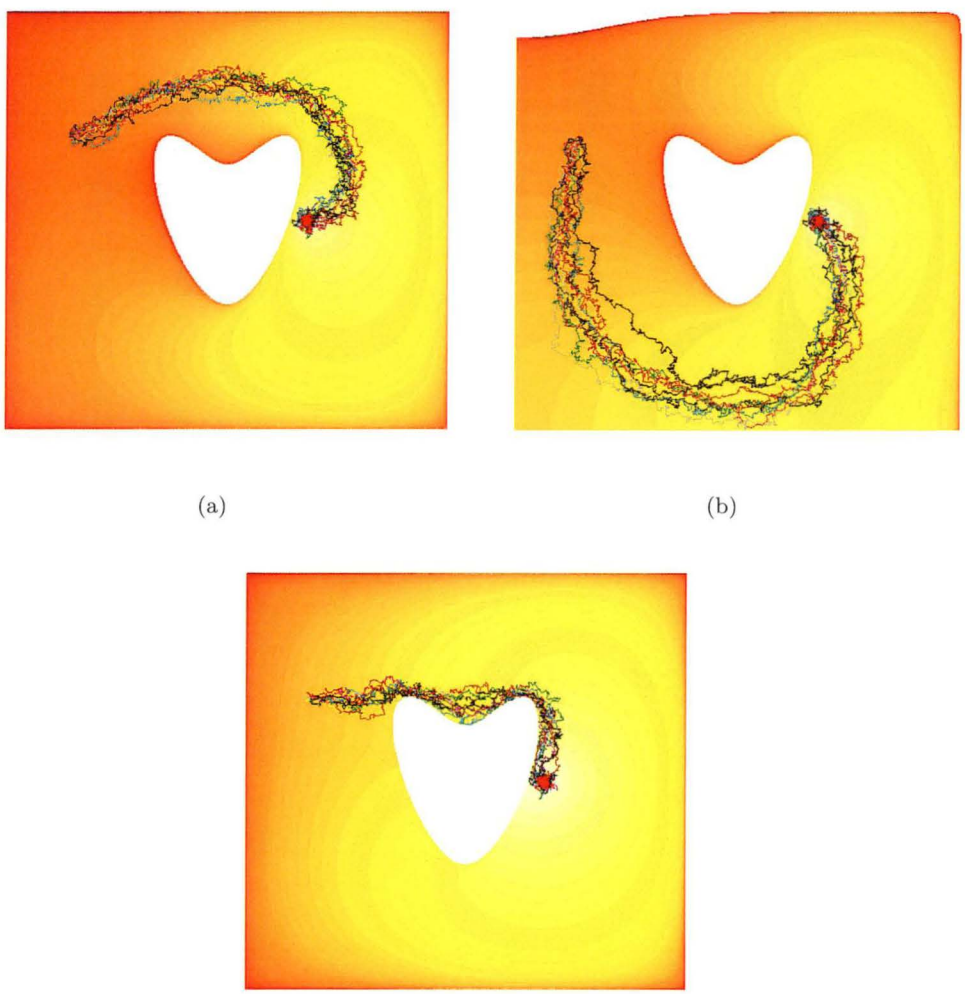
The Dirichlet boundary condition causes the boundary to be cold and therefore to repel a moving animal. When the landmass has this condition, the animal does not move inshore (Figure 7.6 a and b) whereas when the boundary is Neumann it does closely approach the coast (Figure 7.6 c). When the grid boundary has the Neumann condition the animal is able to approach and when this is coupled with a repelling Dirichlet land boundary it is pushed outwards towards the grid boundary (Figure 7.6b); clearly, an undesirable combination. A Dirichlet grid boundary seems best. Whether this is used with a Dirichlet or a Neumann land boundary will depend on the characteristics of the species under investigation - some prefer inshore areas, others avoid them.

In each simulation, all of the animals take the same pathway around the landmass, moving either to the north or to the south. One can imagine examples in which this might be desirable as some animals do take ancestral pathways around landmasses. However, one would usually prefer that a group of individuals split, randomly, when moving around obstacles.

This illustrates the importance of choosing a sensible grid. In this simulation, the land dominates the simulated landscape. Land and grid boundaries are close enough for their boundary effects to combine in their influence on the advection flows around the landmass. There is little space between them for positioning the starting point for the moving animals. This point will be positioned in either a northward or a southward flow, causing all animals to travel in the same direction around the landmass.

By drawing the grid more widely (Figure 7.2) the starting point can be in an area of westerly flow so that as the animals approach land their positions diverge. When they reach land, some will arrive in a position of northward flow and others in southward flow (Figure 7.7c). Those simulations that used a Dirichlet land boundary, which results in a divergent advection field to the northwest of the landmass, require an even larger grid (Figures 7.7a and b).

Another solution is to randomly scatter the animals about the starting point at the beginning of the simulation.



(a) (b)

(c)

Figure 7.6: Paths taken by ten marine animals moving towards a target on the other side of a landmass that dominates the landscape. Boundary conditions were (a) Dirichlet land and grid; (b) Dirichlet land and Neumann grid; and (c) Neumann land and Dirichlet grid. Colours indicate temperature on a log scale (hottest cells are white, coolest red).

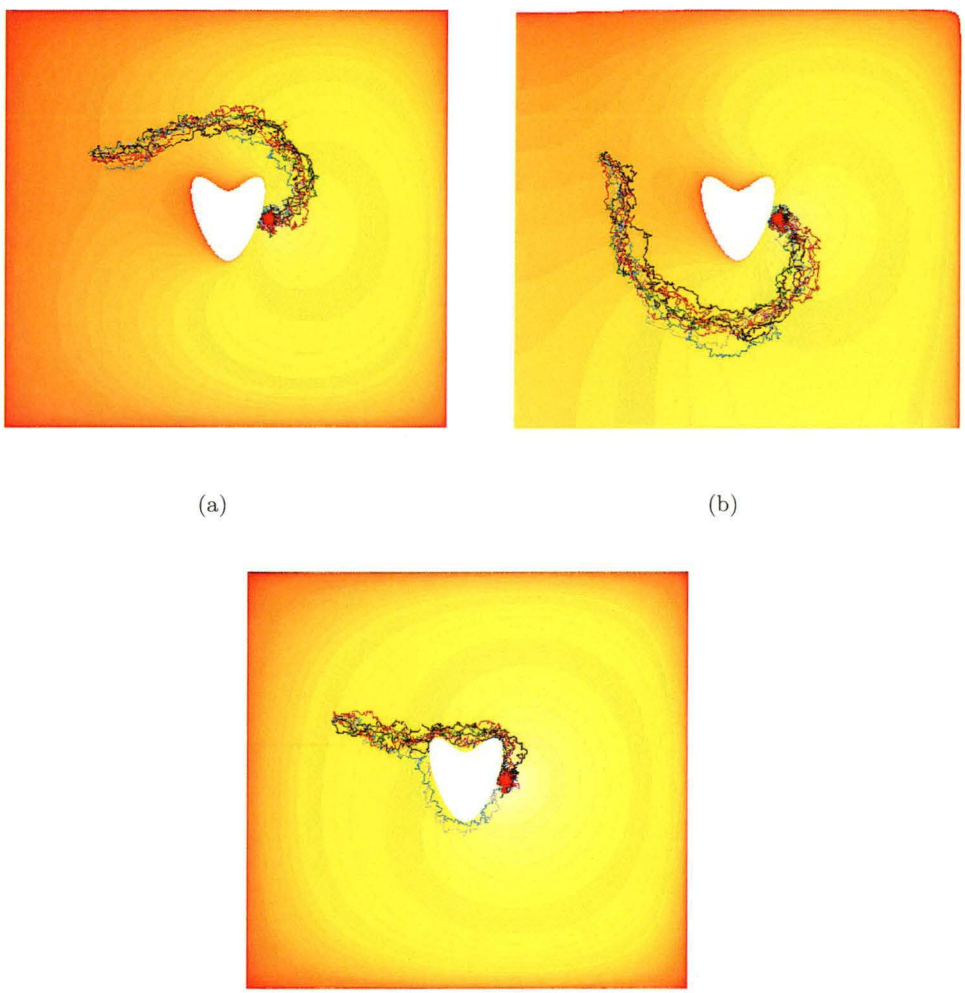
Wherever Laplace's equation is used in this way, care will have to be taken to ensure that undesirable conditions are not imposed on the moving animals.

7.4.2 South Australian sharks

If ten modelled sharks follow a compass bearing (given by the advection shown in Figure 7.4) from a position to the west, directly towards the Neptune Islands, they move into Coffin bay (Figure 7.3). Doing nothing when encountering land results in entrapment (Figure 7.8a). Reversing the direction of travel simply causes the animal to swim in and out of the bay repeatedly (Figure 7.8b), although one animal does escape by random chance. Choosing a random direction in which to swim succeeds only when the simulation unrealistically causes the animal to leap over the narrow peninsula (Figure 7.8c).

When the movement of the ten sharks was directed by an advection flow calculated using Laplace's equation (Figure 7.5) for a heat source located in the middle of the Neptune archipelargo, the sharks all avoid Coffin bay altogether (Figure 7.9). This is true even when the land has the Neumann boundary condition, because the temperature gradient emanating from the target cell draws the sharks further out from the coast. Once again, the Dirichlet land boundary coupled with the Neumann grid boundary seems a poor choice, leading to implausible behaviour (Figure 7.9b).

The variance σ^2 of the random walk affects the step sizes taken by the moving animals. The smooth tracks resulting from the assumption $\sigma^2 = 1^2$ appear unrealistic (7.9). If the variance is large, animals are more likely to step across narrow peninsulas or small islands - with $\sigma^2 = 3^2$ the Neptune islands are obscured by the paths of sharks stepping across them (Figure 7.10). A Dirichlet land boundary can partially help to prevent this by keeping the animals away from some land areas (Figure 7.10a and b), however this solution would not suit animals that do move close inshore, and does not work for the islands at the target location.

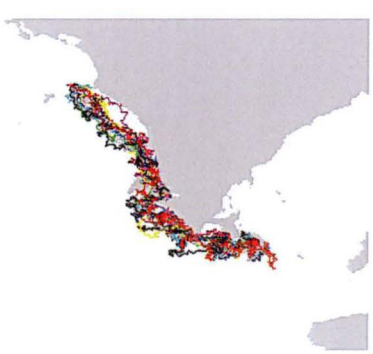


(a) (b)

(c)

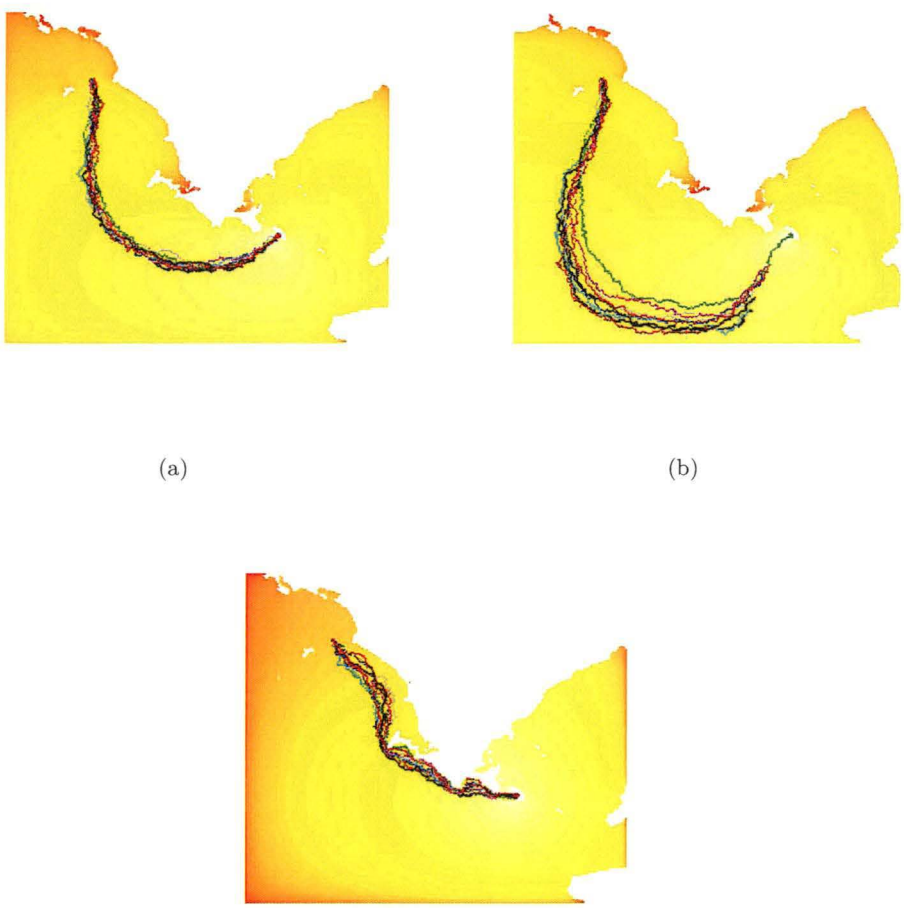
Figure 7.7: Paths taken by ten marine animals moving towards a target on the other side of a landmass. White indicates land, colours indicate temperature on a log scale (hottest cells are white, coolest red). Boundary conditions were (a) Dirichlet land and grid; (b) Dirichlet land and Neumann grid; and (c) Neumann land and Dirichlet grid.

(a) (b)



(c)

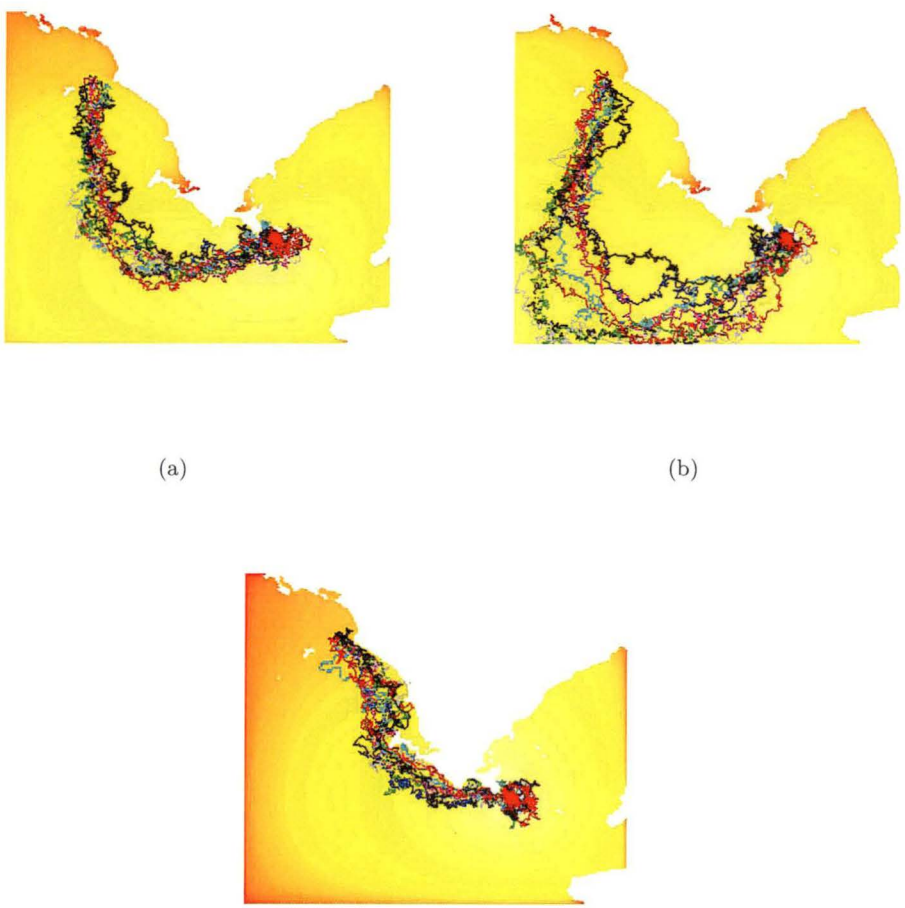
Figure 7.8: Paths taken by ten simulated white sharks following a simple random walk, advected directly towards a target location in the Neptune islands, along the South Australian coast. When coming across land the animals' preferred direction (a) does not change; (b) reverses; or (c) is chosen randomly. Land is coloured grey and the ocean is white.



(a) (b)

(c)

Figure 7.9: Paths taken by ten simulated marine animals moving along the South Australian coast towards the Neptune Islands. Boundary conditions were (a) Dirichlet land and grid; (b) Neumann land and grid; and (c) Dirichlet land and Neumann grid. The random walk used a variance of $\sigma^2 = 1^2$. Colours indicate temperature on a log scale (hottest cells are white, coolest red).



(a) (b)

(c)

Figure 7.10: Paths taken by ten simulated marine animals moving along the South Australian coast towards the Neptune Island. Boundary conditions were (a) Dirichlet land and grid; (b) Neumann land and grid; and (c) Dirichlet land and Neumann grid. The random walk used a variance of $\sigma^2 = 3^2$. Colours indicate temperature on a log scale (hottest cells are white, coolest red).

7.5 Discussion

The problem of simulating the ability of moving animals to avoid entrapment in semi-enclosed parts of the landscape, to move around obstacles, and not to step across narrow barriers, has received little attention in the literature. The solutions employed by investigators are often *ad hoc* - need exists for strategies that are simple to implement, and that reproduce the observed behaviour of animals, while not necessarily accurately reproducing the strategies that they are using. This chapter proposes one solution to the problems of avoiding obstacles and avoiding entrapment in semi-enclosed areas, using Laplace's equation to calculate advection fields that flow around obstacles and past semi-enclosed areas. This is not considered to be the definitive solution to these problems, indeed it does little to prevent simulated animals stepping across narrow barriers. This solution can lead to undesired flows such as the divergence to the northwest generated in Figure 7.1a and c. However, this solution is simple to understand, easily generalizable, and effectively guides simulated moving objects around barriers towards a target. It is hoped that this work will at least serve as a starting point, and a catalyst, for discussion concerning these neglected problems.

This chapter examines gradual replacement of one movement behaviour by another; no scenario is examined in which a single advection field is used, having a time variant co-efficient. Such a scenario would represent the weakening of some behaviour in favour of purely random motion (or the strengthening of some behaviour, away from purely random motion). The purpose of this thesis to develop methods for examining movement behaviour, it does not consider cases of purely random movement, unmodified by deliberate behavioural directives.

CHAPTER 8

Application of movement model and model comparison to tracking data for white sharks in South Australia

8.1 Abstract

The state space movement modelling framework developed in this thesis was applied to observed tracks for white sharks in South Australia. The DIC statistic was used for model comparison. The hypothesis was examined that the sharks were navigating by following the 80m isobath (line of bathymtry). Alternative hypotheses, presented as foils to the isobath-following hypothesis, were that sharks remained 50m from shore, and that sharks travel in certain compass directions. In simulation, the DIC was able to successfully discern that the isobath hypothesis had been used to simulate tracks, even though the alternative hypotheses lead to very similar advection fields.

When all advection fields were included in a single movement model, the 95% credibility interval for the coefficient for the isobath-following field was the only one that did not include zero. This indicates that the model can be used in two ways to make inference about movement: using DIC applied to competing movement models, and by including competing movement hypotheses in a single model.

Application of these methods to the observed tracks showed little support for the isobath-following hypothesis. Most support was shown for the hypothesis that the animals follow compass bearings - the least explanatory of the hypotheses examined.

The movements of the sharks appeared to be more complex than any of the proposed hypotheses, in particular, behaviour appears to switch between travelling and foraging states. Future applications of this model will consider the use of mixtures of normal distributions in order to model such behaviour-switching.

8.2 Introduction

Conventional tagging methodology has yielded a great deal of knowledge regarding the broad movements and population dynamics of sharks (Kohler & Turner, 2001). Now, newer satellite tracking provides an opportunity for evaluating hypotheses regarding animal navigation (Alerstam, 2003). The movements of animals that undertake long ocean crossings have been shown to be consistent with navigation using the position of the sun and stars (Leis & Carson-Ewart, 2003; Bonfil *et al.* , 2005), light polarization patterns, odours, wind and water currents, temperature gradients and salinity (Quinn, 1994), the earth's main magnetic field (Boles & Lohmann, 2003; Montgomery & Walker, 2001; Cain *et al.* , 2005) and magnetic anomalies (Klimley, 1993).

Elasmobranch movements have been attributed to responses to a large number of internal and external cues, including the general condition of the shark, water currents, light levels, time of year, geographic location, water masses, oxygen levels, and in particular, presence or absence of prey, ambient water temperature, bottom type and magnetic gradient (Sundstrom *et al.* , 2001). Individual white sharks *Carcharodon carcharias* undertaking movements between regularly frequented areas have been shown to follow similar paths, termed migration corridors (Weng *et al.* , 2007) or common highways (Bruce *et al.* , 2006). White sharks have been recorded moving considerable distances across oceans, from the south coast of South Africa to western Australia and back (Bonfil *et al.* , 2005), from southern Australia to New Zealand (Bruce *et al.* , 2006), from New Zealand to the Australian Great Barrier Reef (NIWA press release, 24 December 2007) and from the Californian coast to the Hawaiian islands. The means by which white sharks navigate between these areas is unknown, although celestial cues, such as the position of the sun or star constellations, have been suggested (Bonfil *et al.* , 2005).

White sharks, like many other shark species, show 'yo-yoing' oscillatory diving while travelling, in that they repeatedly swim from the surface to depth and back again (Klimley *et al.* , 2002; Bruce *et al.* , 2006; Carey & Scharold, 1990). The purpose

of these oscillations is unknown, but this behaviour would allow sharks, at least in waters shallower than their maximum diving range (at least 450m, Boustany *et al.* (2002)), to assess the depth of the water through which they are travelling. This would be true even if depth sounding were not the primary purpose of the oscillatory movements (Bruce *et al.* , 2006).

White sharks may use a range of navigational strategies and cues, and may use different strategies for long ocean crossings than for coastal movements. Bruce *et al.* (2006) have speculated that in coastal areas of South Australia white sharks may navigate by following lines of bathymetry. In this chapter, satellite tracks collected from three white sharks tagged off the South Australian coast are used to investigate this suggestion. A fourth track is used to illustrate the importance of selecting a smooth advection field to represent a hypothesis of movement.

Previous chapters presented a state space model capable of simulating a wide range of movement behaviours. Simulation tests showed that the model can correctly estimate the coefficients of the advection fields used to simulate a track (Chapter 3). Although precision parameters for the errors in the track and the observations were not well estimated, this did not affect the ability of the model comparison measures, (the Bayes factor and the deviance information criterion DIC) from correctly identifying the candidate behaviour that was used to simulate the track. It was shown (Chapter 5) that the DIC performs at least as well as the more computationally intensive Bayes factor, and performs better when a highly diffuse prior is used for the precision parameters.

This chapter applies the movement model developed in earlier chapters to tracking data collected from white sharks in South Australia and uses the DIC to gauge which of several competing hypotheses regarding navigational behaviour is given most support by the data. The focus is on the theory that white sharks navigate by following a line of bathymetry (an isobath). The opposing hypotheses against which this is tested were chosen for their similarity to the isobath-following advection field. Simulation testing was used to ascertain that when tracks are simulated using the isobath-following hypothesis, then the modelling framework used here is able to distinguish this hypothesis from the similar, alternative hypotheses.

8.3 Methods

8.3.1 Data

Tracks from four white sharks moving off the coast of South Australia (Figure 8.1) were obtained as part of an on-going investigation into the movements of white sharks in Australian waters, Bruce *pers commn.* Bruce *et al.* (2006) describes the methodology and initial findings of that study. All four sharks were tagged during November 2004 and were tracked until February 2005 using Argos satellite tags. Only those sections of the track in which the shark moved within this coastal region of South Australian were used, because this is the area in which navigation by depth is suggested to be evident (Barry Bruce, *pers comm.*). Sharks 1 and 3 traversed the region and then turned sharply and began to retrace their steps. In these cases, the track was truncated at the turning point.

Bathymetry data were obtained from Geosciences Australia (Petkovic & Buchanan, 2002) at a resolution of 0.01 degrees of latitude by 0.012 degrees of longitude. This gives a roughly 1 × 1km grid. This application assumes that movement occurs on a Euclidean plane, rather than a sphere - a reasonable assumption on the relatively small scale of this investigation.

A smooth advection field is necessary because the Metropolis-Hastings step, used to make draws of the path (true) locations of the tracked animal from the posterior, requires that advection be locally constant (see Section 3.5). Smoothing the bathymetry correspondingly smooths the advection field that represents navigation by following an isobath.

The bathymetry data were smoothed using six passes of a 25×25 cell block smoother. Each cell that was more than 12 cells away from the edge of the landscape grid was considered to be the central cell of a 25 × 25 cell block, and its depth was replaced by the average depth of the surrounding block. Cells that were within 12 rows or columns of the grid boundary were not smoothed, as sharks did not approach these boundaries. The size of the block, and the number of passes by the smoother, were chosen so that the resulting advection field for isobath following appeared smooth.

Nevertheless, the resulting advection field is not perfectly smooth (Figure 8.2a). However, this is probably realistic because a white shark's perception of the bathymetric landscape is unlikely to be perfect. The sharks' oscillatory diving would

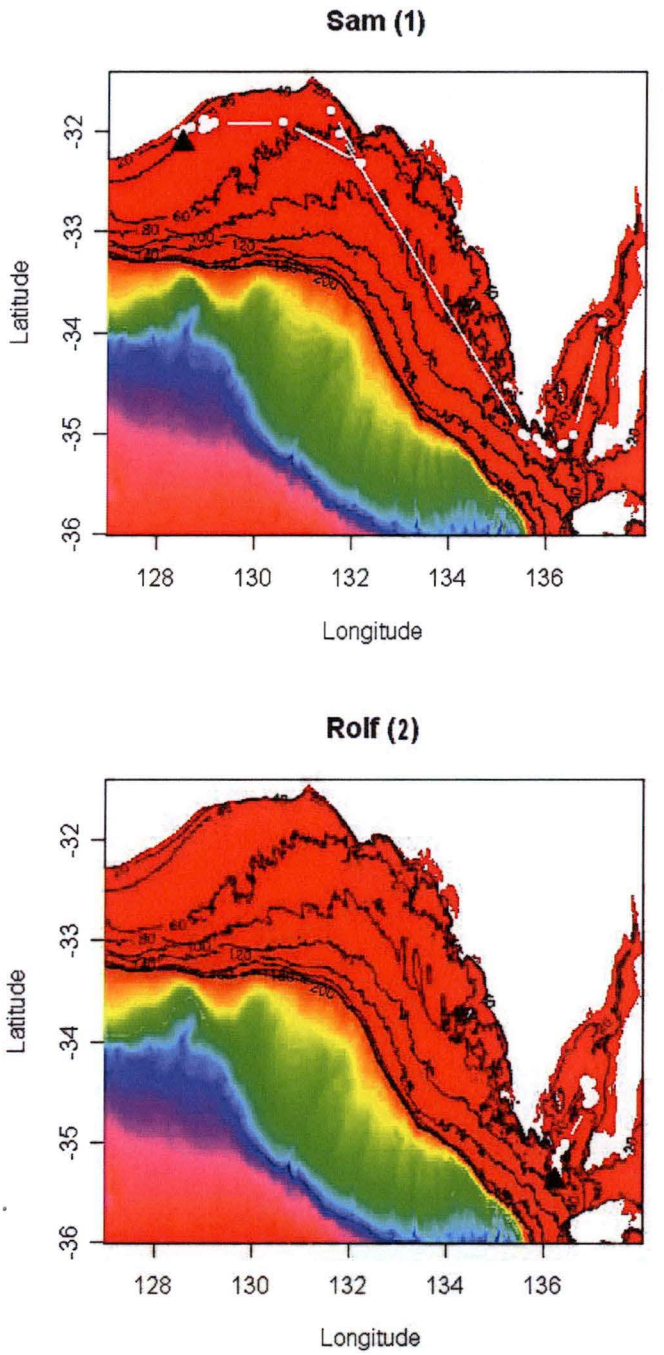


Figure 8.1: Sections of observed tracks for four white sharks (identified by name and number) tagged off South Australia. The track is white and the first location is a black triangle. Bathymetry is indicated by shading and by contours, which are shown every 20m up to a depth of 200m.

Continued overleaf

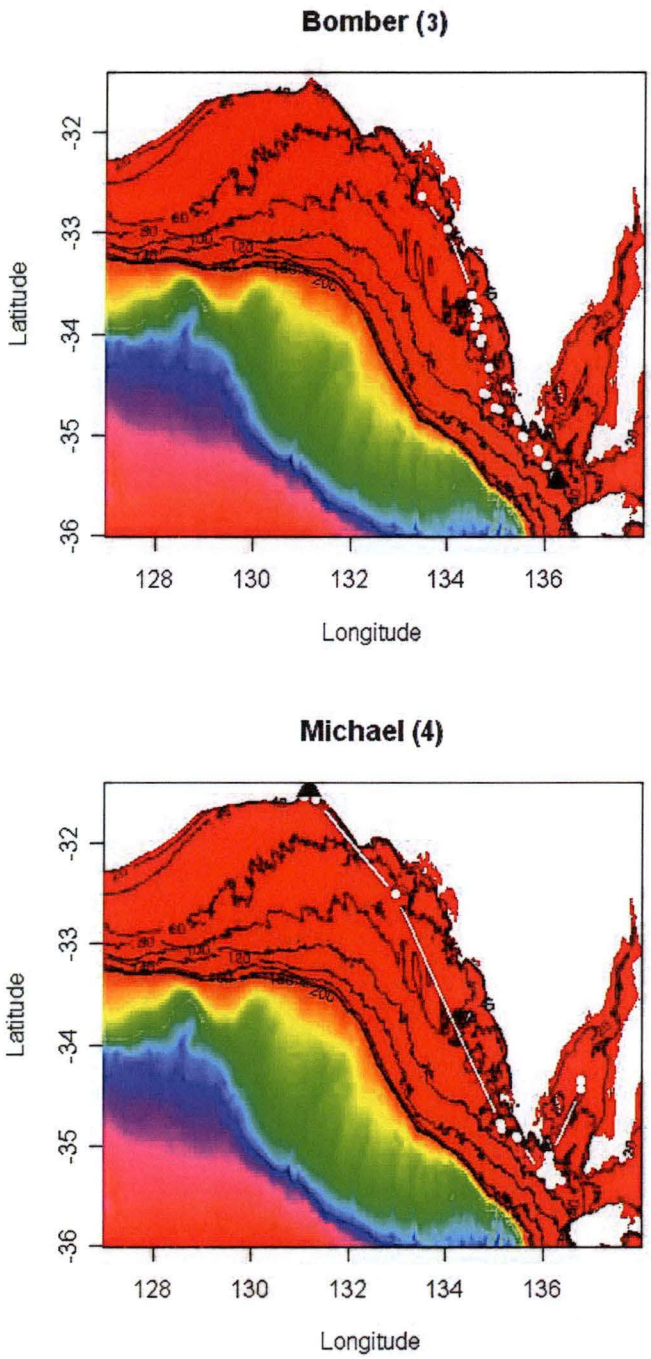


Figure 8.1. *Continued from previous page.* Sections of observed tracks for four white sharks (identified by name and number) tagged off South Australia. The track is white and the first location is a black triangle. Bathymetry is indicated by shading and by contours, which are shown every 20m up to a depth of 200m.

provide a sampling of depths at intervals. If sharks do use bathymetry to navigate, they would be using an approximation to the actual bathymetric landscape. The smoothing routine applied here might mimic the imperfect perception of bathymetric gradients that might be used by a real shark in navigating towards a goal.

8.3.2 Simulated tracks

The movement model applied in this chapter has already been simulation tested (Chapter 3). However, aspects of the problem presented in the present chapter are novel and might result in difficulties in identification of the hypothesis that best supports the data. Novel aspects include the presence of land, advection fields that are more likely to violate the assumption of locally constant advection (see Section 3.5), and large variance in the duration of time steps. In addition, it is unknown whether the model can distinguish between the relatively similar advection fields used in this chapter (and described in the following sections).

Therefore, for each of the four observed tracks, an additional four tracks were simulated. These simulations used the observed starting location x_1, the observed time steps Δ_k, observed number of locations N (with the addition of latent steps, as described in Section 8.3.5 below), and observed overall direction of travel (eastward or westward).

The simulated location of the track y_k at time step k, $k \in O$ is a function of the true location x_k of the animal, and normally distributed error ϵ_k. This was given by equation 3.1, and is repeated here

$$y_k = x_k + \epsilon_k \quad , k \in O. \tag{8.1}$$

Some true locations may be unobserved (latent) so that O is a subset of the time steps $k = 1, \ldots, N$.

Four paths were simulated by starting each animal at its observed starting location x_1, and then advancing the animal by the observed, plus latent, number of steps using equation 3.3. This equation is repeated here, and for simplicity of pre-

sentation no subscript is used to denote the individual shark

$$x_{k+1} = x_k + \Delta_k \left[v_k + (X_k\, \beta)^T\right] + \xi_k, \qquad k = 1 \ldots N. \tag{8.2}$$

The inertial term, velocity v_k, which represents a tendency for the individual to show directional persistence, is given by a random walk (equation 3.4, repeated below)

$$v_{k+1} = v_k + \zeta_k. \tag{8.3}$$

The errors in the track ϵ_k were drawn randomly from a normal distribution with mean zero and precisions τ^y (see equation 3.2, repeated below)

$$\epsilon_k \sim \mathrm{N}\left(0, (\tau^y)^{-1}\right) \qquad k = 1 \in O. \tag{8.4}$$

The errors in the path ξ_k and velocity ζ_k were drawn randomly from normal distributions with mean zero and precisions scaled by the duration of the time step (equation 3.5, repeated below)

$$\xi_k \sim \mathrm{N}\left(0, \Delta_k\, (\tau^x)^{-1}\right) \qquad k = 1 \ldots (N-1) \tag{8.5}$$

$$\zeta_k \sim \mathrm{N}\left(0, \Delta_k\, (\tau^v)^{-1}\right) \qquad k = 1 \ldots (N-2). \tag{8.6}$$

The advection field X used to simulate the track pertained to navigation by following an isobath (the 'Depth' field described in Section 8.3.3). Shark 3 (Bomber) moved westwards whereas all other sharks moved eastwards so that a different variant of this field was used for shark 3.

The parameter values used for the simulation are shown in Table 8.1.

Table 8.1: Parameter values and settings used when simulating four shark tracks. The starting locations, time steps and number of locations were those observed when tracking four white sharks in South Australia.

Symbol	Description	Value
$N-n$	Number of observed locations (sharks 1-4)	$39, 38, 24, 44$
n	Number of latent locations (sharks 1-4)	$14, 0, 1, 26$
m	Number of advection fields	1 or 3
x_1	Initial location, Shark 1	$(128.510, -32.123)$
	Shark 2	$(136.250, -35.386)$
	Shark 3	$(136.277, -35.456)$
	Shark 4	$(131.172, -31.442)$
Λ	Step durations	Range: 0.25h to 48h
β	Advection coefficient (for 'Depth')	5
$\tau_1^y = \tau_2^y$	Precision for observations	$(1, 1)$
$\tau_1^x = \tau_2^x$	Precision for path	1
$\tau_1^v = \tau_2^v$	Precision for velocity	0.1^{-1}

Michael: movement close to land

The first observed location for shark 4, Michael, was discarded because it was on land (doubless due to an observation error). The second location was very close to land. The advection field that reflects westward movement, navigating by isobath-following, flows westwards in the region of this observed location. There is very little bathymetric slope in this region so that the westward flow dominates, pushing the simulated track onto the land. Rejection sampling was used to prevent the simulated track moving onto land, but for 2 steps a non-land location could not be found after 10 000 draws. In these cases, the most recently drawn non-land location was accepted.

8.3.3 Advection fields

Three alternative advection fields were generated, each corresponding to a different hypothesis regarding how white sharks might be navigating. The means of turning these hypotheses into advection fields are described in the following sections.

Over most of the region considered, the ocean lies to the south of the land, therefore, over land, all advection fields specify a southerly flow in order to deflect sharks that reach land, back towards the water.

Depth: isobath following

The hypothesis was examined that white sharks, when travelling eastward or westward along the south coast of Australia, prefer to travel along the 80m line of bathymetry (isobath) (Bruce *et al.* , 2006), which they use as a navigational aid. Advection with respect to longitude (the east-west direction) was given by a constant, the sign of which determines whether movement will be eastward or westward. Advection with respect to latitude (the north-south direction) was calculated from smoothed bathymetric data for the South Australian region. The resulting advection field flows eastward or westward with a bias towards the 80m depth contour.

First, for every cell in the grid of latitude h and longitude j the absolute value A_{hj} of the difference between the depth D_{hj} of that cell and the preferred depth (80m) was calculated

$$A_{hj} = \mid D_{hj} - 80 \mid . \tag{8.7}$$

Advection with respect to longitude X_1 is given by a constant W, normalized by L_{hj}, the length of the advection vector for the hth row and jth columns, so that the result is a unit vector

$$(X_1)_{hj} = \frac{W}{L_{hj}}. \tag{8.8}$$

Advection with respect to latitude X_2 is given by the normalized first differences in A_{hj} for the columns adjacent to column j, normalized by L_{hj}

$$(X_2)_{hj} = \frac{(A_{hj-1} - A_{hj+1})}{L_{hj}}. \tag{8.9}$$

The normalising factor L_{hj} is the length of the vector given by the numerators above

$$L_{hj} = \sqrt{W^2 + (A_{hj-1} - A_{hj+1})^2} \tag{8.10}$$

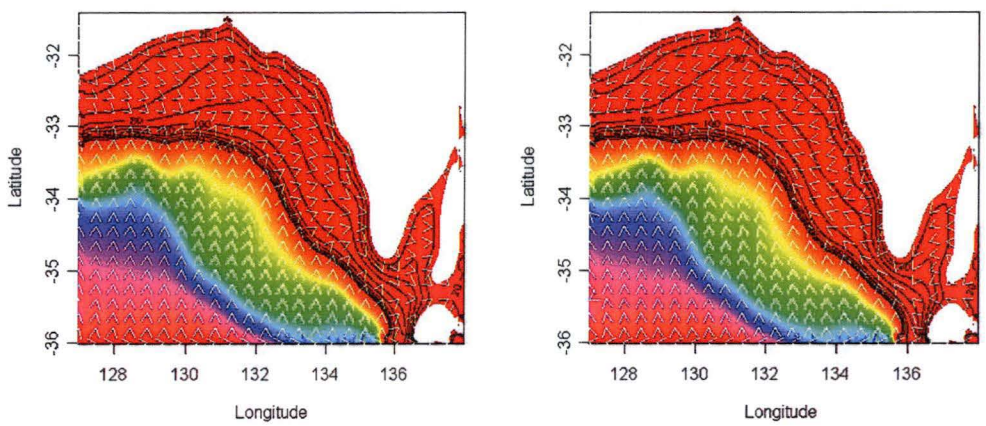
so that $[(X_1)_{hj}, (X_2)_{hj}]$ is a unit vector. Unit vectors are used so that the speed of movement of the shark is governed by the advection coefficients β and the precision parameters for the path and for velocity, and not by the bathymetric slope.

This results in two advection fields, one in which W has positive sign, so that sharks move in an easterly direction (8.2a) and another in which W is negative so that sharks move westwards (8.2b). The advection fields that result from unsmoothed bathymetric data is shown in Figure 8.3, for contrast.

Note that the value of the constant W affects the extent to which eastward or westward advection dominates the north-south direction given by bathymetry. A value of $W = 4$ was used here. Larger values resulted in excessively weak iso-bath directed flow, and smaller values resulted in sharp changes in advective direction across the preferred isobath. The resulting zig-zag pattern of a simulated track is shown in Section 8.4.1 and the difficulties that such sharp changes present to the MCMC routine are discussed in Section 8.4.2.

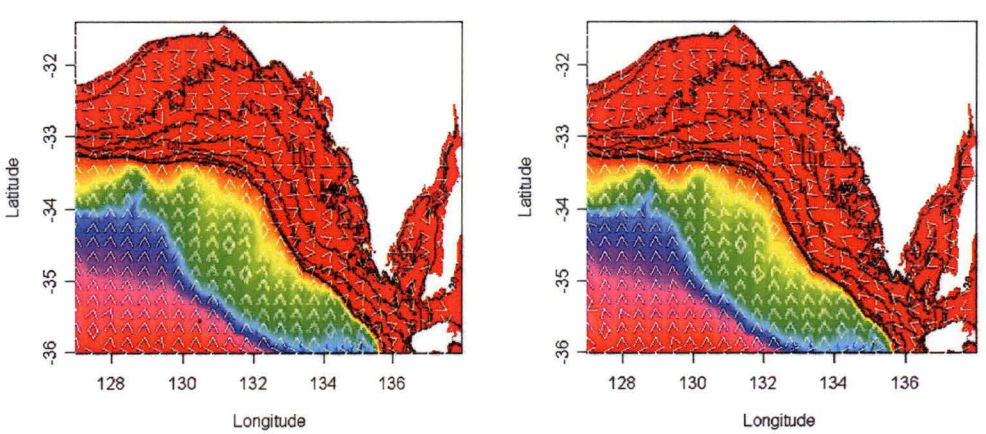
Distance: coastline following

The 'Depth' hypothesis is the primary focus of this investigation, but alternatives are required against which to test that hypothesis. A closely related hypothesis for this region where isobaths parallel the coast, is that sharks follow the coastline, remaining a constant distance away from it. A distance of 50km from the shore was chosen after examination of the observed tracks, so as to reflect the modal distance



(a) (b)

Figure 8.2: Smoothed bathymetric data were used to calculate advection fields (white arrows) representing navigation by following the 80m isobath and a preference for travel (a) eastwards, or (b) westwards. Bathymetry is indicated by shading and by contours, which are shown every 20m up to a depth of 200m.



(a) (b)

Figure 8.3: Unsmoothed bathymetric data results in somewhat erratic advection fields (white arrows) representing navigation by following the 80m isobath and a preference for travel (a) eastwards, or (b) westwards. Bathymetry is indicated by shading and by contours, which are shown every 20m up to a depth of 200m.

kept by the observed sharks (recall that this hypothesis is included as a foil for the 'Depth' hypothesis, rather than as a proposed navigational strategy).

An advection field reflecting the hypothesis of navigation by coastline following, was calculated in the same way as that for isobath following, except that the depth of each cell was replaced by its distance to the nearest coastal cell. Coastal cells were defined as those that have at least one land neighbour and at least one water neighbour. Non-coastal cells are either water (depth greater than zero), or land. For each water cell, the distance to each of the land cells was calculated using the Haversine formula, which assumes a spherical earth, to calculate the distance between two latitude-longitude co-ordinates (Sinnott, 2004). The smallest of these distances was retained d. The smoothed bathymetry data were used to define the coastline, in this instance smoothing may be regarded as equivalent to the natural integration that might occur due to imperfect sensing of the location of the coastline.

For each cell in row h and column j the absolute value $\widetilde{A}_{hj}$ of the difference between the distance to the nearest landmass d and the preferred distance (50km) is given by

$$\widetilde{A}_{hj} = \mid d_{hj} - 50 \mid . \tag{8.11}$$

Advection with respect to longitude X_1 is again given by a constant $\widetilde{W}$, normalized by $\widetilde{L}_{hj}$, the length of the advection vector, so that the result is a unit vector

$$(X_1)_{hj} = \frac{W}{\widetilde{L}_{hj}}. \tag{8.12}$$

Advection with respect to latitude X_2 is again given by the normalized first differences in $\widetilde{A}_{hj}$ for the columns adjacent to column j, normalized by $\widetilde{L}_{hj}$

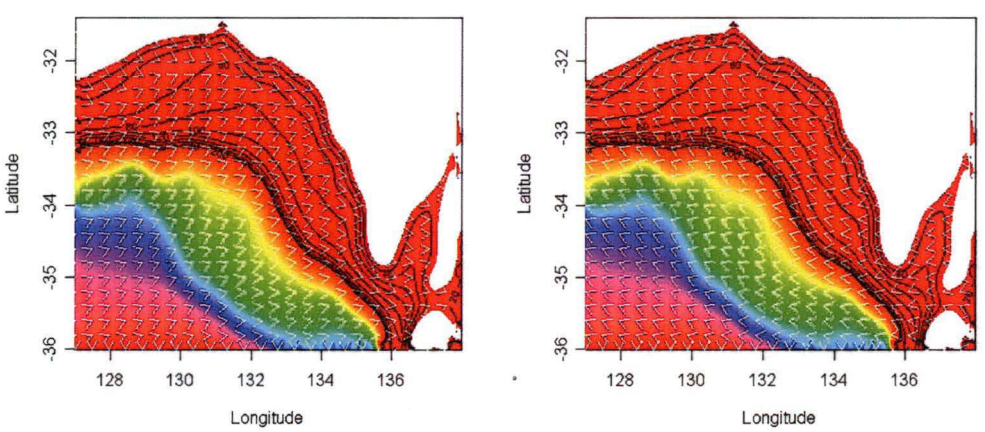
$$(X_2)_{hj} = \frac{\left(\widetilde{A}_{h\,j-1} - \widetilde{A}_{h\,j+1}\right)}{\widetilde{L}_{hj}}. \tag{8.13}$$

The normalising factor $\widetilde{L}_{hj}$ is the length of the vector given by the numerators above

$$\widetilde{L}_{hj} = \sqrt{\widetilde{W}^2 + \left(\widetilde{A}_{h\,j-1} - \widetilde{A}_{h\,j+1}\right)^2} \tag{8.14}$$

so that $[(X_1)_{hj}, (X_2)_{hj}]$ is a unit vector, as explained in Section 8.3.3. For notational clarity, subscripts indexing advection field (such as 'Depth' or 'Distance') are omitted from the symbols X_1 and X_2.

Again, both positive and negative signed values are used for the constant $\widetilde{W}$ so that two advection fields are derived, one for sharks that move in an easterly direction (Figure 8.4a) and another for sharks that move westwards (Figure 8.4b).



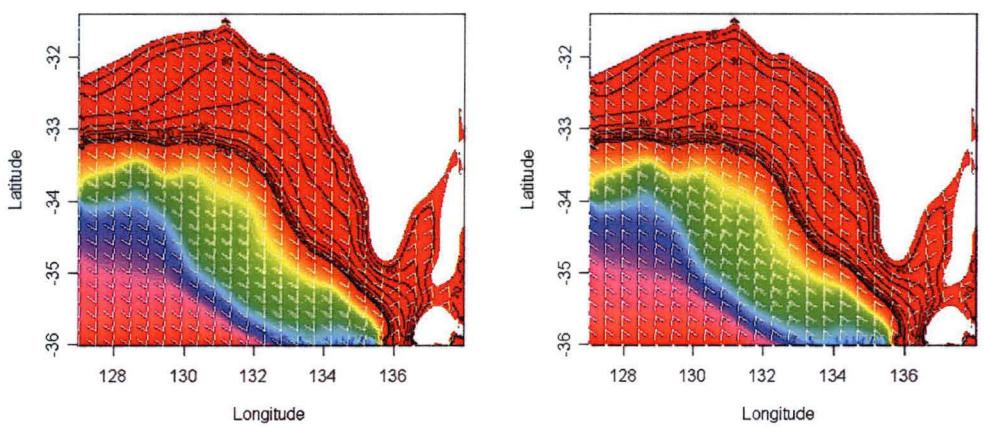
(a) (b)

Figure 8.4: White arrows indicate advection fields representing a preference for remaining 50km from the coast and for moving (a) eastwards, and (b) westwards. Bathymetry is indicated by shading and by contours, which are shown every 20m up to a depth of 200m.

***Direction*: preferred direction**

A simple alternative hypothesis, which reflects a lack of understanding regarding how sharks navigate, is that sharks move due southwest or northeast when in Spencer Gulf, and due northwest or southeast when westwards of Spencer Gulf (Figure 8.5).

The corresponding advection fields are shown in Figure 8.5.



(a) (b)

Figure 8.5: White arrows indicate advection fields representing a preference for travel (a) southeasterly in Spencer Gulf and northeasterly outside of it, and (b) northeasterly in Spencer Gulf and southeasterly outside of it. Bathymetry is indicated by shading and by contours, which are shown every 20m up to a depth of 200m.

8.3.4 Movement model and estimation

The state space model for movement described in Chapter 3 is used in the present chapter. Observation errors were estimated separately for latitude and longitude due to the known error structure of satellite tracking data (described in Section 2.5.5). The Argos error classification was ignored in this first application to observed tracks. The deviance information criterion DIC was used to evaluate the support given by the data to alternative choices of advection field.

For each tracked animal (for both the observed and simulated tracks), each of four models were applied. Each model used a different (set of) advection field(s):
1. Depth: the 80m contour is followed (Figure 8.2),

2. Distance: the shark follows the coastline at a distance of 50km (Figure 8.4),
3. Direction: movement is due NW-SE or NE-SW (Figure 8.5),
4. All: all three of the above methods are used by the shark.

The fourth model, which combines all advection fields, was introduced in order to investigate the ability of the model to eliminate incorrect advection fields by estimating advection coefficient values close to zero. This ability can be evaluated by applying the model to the tracks that were simulated using only the 'Depth' advection field.

The priors used during estimation are shown in Table 8.2. Note that the prior for the precision for velocity τ^v is larger than that used in previous chapters (0.1^{-1} instead of 0.01^{-1}). This allows for the greater variation in speed of travel by real sharks than that simulated in previous chapters. Real sharks travel at speed between areas and are then resident for a period, which is equivalent to slowing down.

A larger variance is used for the prior for the advection coefficients β (100 instead of 1) which reflects the larger values used for β (5 instead of 1). Larger β values more closely approximated the overall observed speed (and therefore distance covered) of the sharks.

Table 8.2: Prior distributions used in the movement model.

Symbol	Description		Value
x_0	Mean initial location,	Shark 1	$(128.510, -32.123)$
		Shark 2	$(136.250, -35.386)$
		Shark 3	$(136.277, -35.456)$
		Shark 4	$(131.155, -31.471)$
τ^{x0}	Precision for initial location		1000
v_0	Mean initial velocity		$(0, 0)$
τ^{v0}	Precision for initial velocity		1000
β_0	Mean of multivariate normal prior for β		0 or $(0, 0, 0)$
Σ_0	Variance for prior for β		100 or diag$(100, 100, 100)$
a_1^y, a_2^y, a^x	Scale for gamma priors for τ^y and τ^x		$(1, 1, 1)$
b_1^y, b_2^y, b^x	Rate for gamma priors for τ^y and τ^x		$(1, 1, 1)$
	Mean for gamma priors for τ^x and τ^y		$(1, 1, 1)$
	Variance for gamma priors for τ^x and τ^y		$(1, 1, 1)$
a^v	Scale for gamma prior for τ^v		$(100, 100)$
b^v	Rate for gamma priors for τ^v		$(10, 10)$
	Mean for gamma priors for τ^v		$(0.1, 0.1)$
	Variance for gamma priors for τ^v		$(1, 1)$

8.3.5 Time and latent locations

Time steps were measured in 12 hour units. Latent locations, equally spaced in time, were introduced between any two observations that were more than 48 hours apart. The smallest number of latent locations required to reduce each time step to 48 hours or less, was used.

A large number of observations occurred within seconds or minutes of one another, most occurring in pairs. This probably indicates that the usual time spent at the surface by a shark is roughly twice the average time required to establish a position fix. However, it is possible that locations could occur in clusters, which would give excessive weight to some locations. These repeated measures greatly increased computation time, and caused difficulties with computer memory overflow. To solve this problem, any observed location that occurred within 15 minutes of an earlier observation was excluded. This eliminated between 48% and 52% of the observed locations for each track. The resulting loss of information is discussed in Section 8.5.

8.3.6 DIC and MCMC

The Markov chain Monte Carlo method used to estimate posterior quantities for the movement model was described in Section 3.3.6, and the method of calculating the deviance information criterion DIC from such a set of MCMC draws from the posterior, was described in Section 5.3.5. The MCMC and subsequent DIC calculations were repeated 5 times, in order to calculate a measure of spread for the DIC (see Section 5.3.7). In this case, the standard deviation is reported.

When using the observed track for shark 1, Sam, it was necessary to thin the MCMC chain to achieve apparent convergence. For the models that used the 'Direction' and 'Distance' hypotheses, every 20th draw of the chain was retained; for the 'All' model every 30th, and for the 'Depth' model every 100th draw was retained. Thinning was required due to sharp changes in the advection field in the vicinity of the coastline, this is discussed in Section 8.4.2.

8.4 Results

8.4.1 Simulated tracks

For each observed track, another track was simulated using the observed starting location, number of locations, and time steps. Latent locations were added so that no time step was longer than 48 hours.

The 'Depth' advection field was used to simulate these tracks. This field was generating using a weight $W = 4$, which was found to provide a good compromise between dominance of east-west motion over north-south (isobath following) motion. When larger values are used the east-west motion almost eliminates the isobath-directed motion, and when smaller values are used simulated tracks can converge rapidly on the 80m isobath and then zig-zag across it (e.g. Figure 8.6).

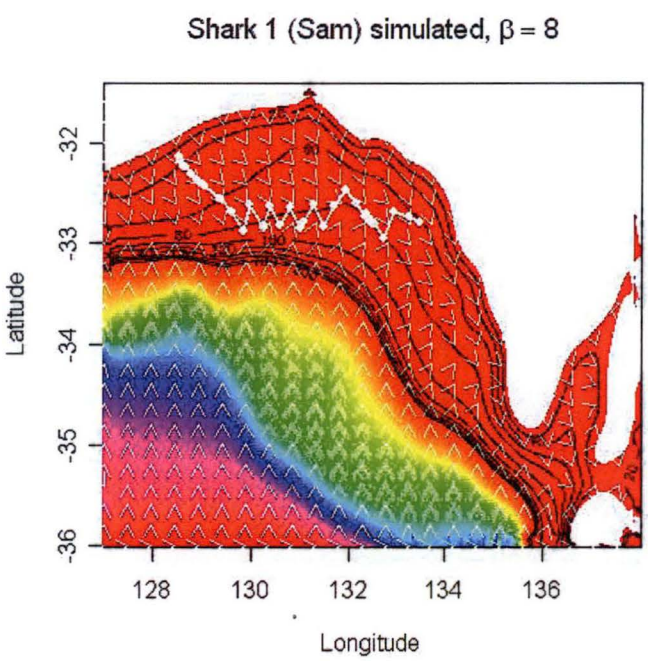


Figure 8.6: Simulated track for shark 1, Sam, using an advection field (white arrows) that reflects a preference for the 80m isobath. A weight of $W = 1$ allows north-south isobath-directed movement to be relatively dominant over east-west movement, resulting in a zig-zag track across the isobath. A relatively large advection coefficient of $\beta = 8$ has been used to make this pattern clearer.

The tracks for sharks 1 and 4 included several latent locations so that the 'observed' subset, when joined by straight lines, gives a misleading impression of the path

taken by the shark (white connected dots in Figure 8.7). Track 2 has no latent locations and track 3 has only one (Table 8.1) so that the path and 'observed' track are indistinguishable (Figure 8.7).

The simulated track for shark 2, Rolf, does not move northwards like its observed counterpart because the advection field based on following an isobath in an eastward direction does not allow for northerly movement into this gulf. The observed track for shark 2 cannot, therefore, be expected to support the hypothesis that the 'Depth' advection field governs movement (Figure 8.1 vs Figure 8.7).

The track for shark 3, Bomber, covers less distance than the corresponding observed track (Figure 8.1 vs Figure 8.7) indicating that the observed shark had a higher overall speed of travel.

8.4.2 Convergence: movement close to land

A southerly advection flow was chosen over land - resulting in a sharp change in the direction of the advection field at the coast. Such a change violates the assumption of locally constant advection that underpins the decision to use the 'conditional' distribution for the path x as its proposal distribution (see Section 3.3.6). This proposal distribution would be the full conditional distribution for the path x, if advection were constant. When advection is not locally constant, this becomes a poor choice.

Most tracks did not stray close to the land so that the sharp change in advective direction at the coast presented no difficulties. However, the observed track for shark 1 has several locations close to land. This was sufficient to result in poor mixing of the MCMC chain for the path x (Figure 8.8a), and a consequent need to thin the chain (Figure 8.8b).

This difficulty is more strongly illustrated by Shark 4, Michael. The first observed location for Michael was on land and subsequent observed locations were closer to land that those of the other sharks (Figure 8.1). As discussed in the Methods section, two simulated locations for this shark were also on land (Figure 8.7) because, in the near-inshore region, the advection field used to simulate the track directed this shark towards land. Because both the observed and simulated tracks include locations in the coastal region, where advection changes direction sharply, both fail to converge when the MCMC chain is not thinned (Figure 8.9). Extensive thinning

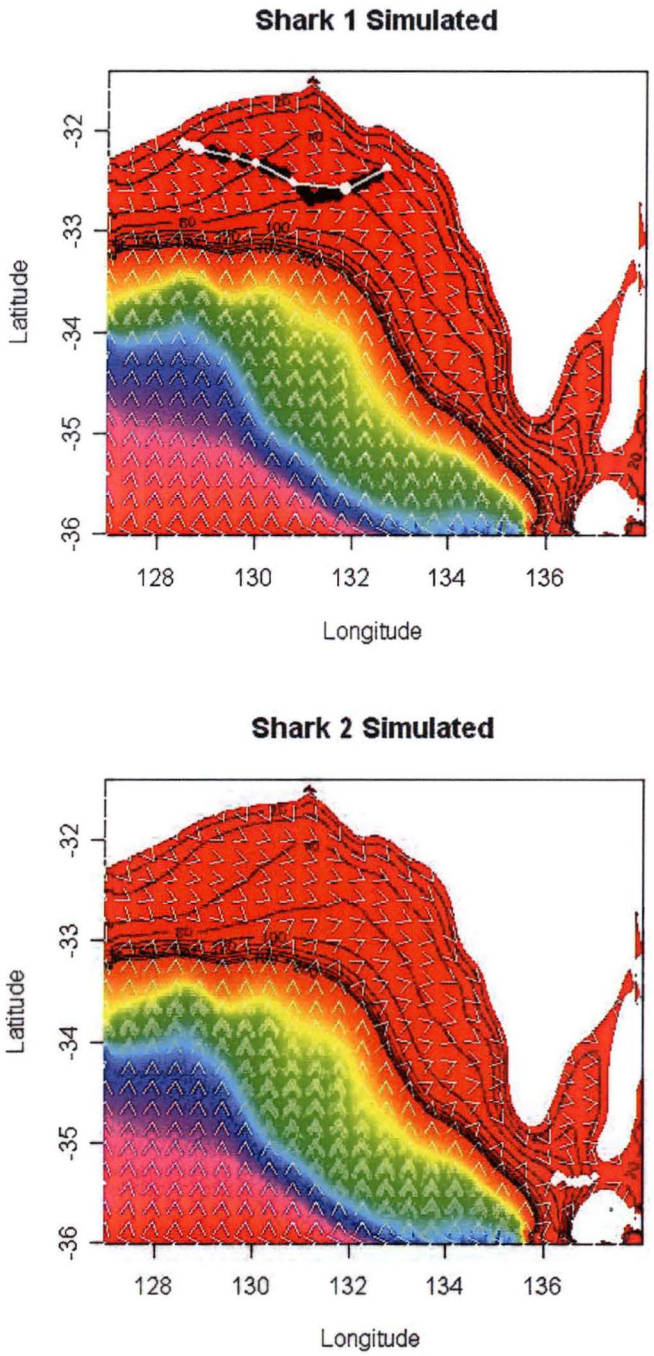


Figure 8.7: Simulated tracks (shown as connected white dots) that used the observed start location, number of locations and time steps observed for four white sharks (identified by name and number) tagged off South Australia. The simulated path is shown as black dots, the 'observed' locations are white dots joined by a white line. The tracks were simulated using the 'Depth' advection field (shown as white arrows). Bathymetry is indicated by shading and by contours, which are shown every 20m up to a depth of 200m.

Continued overleaf

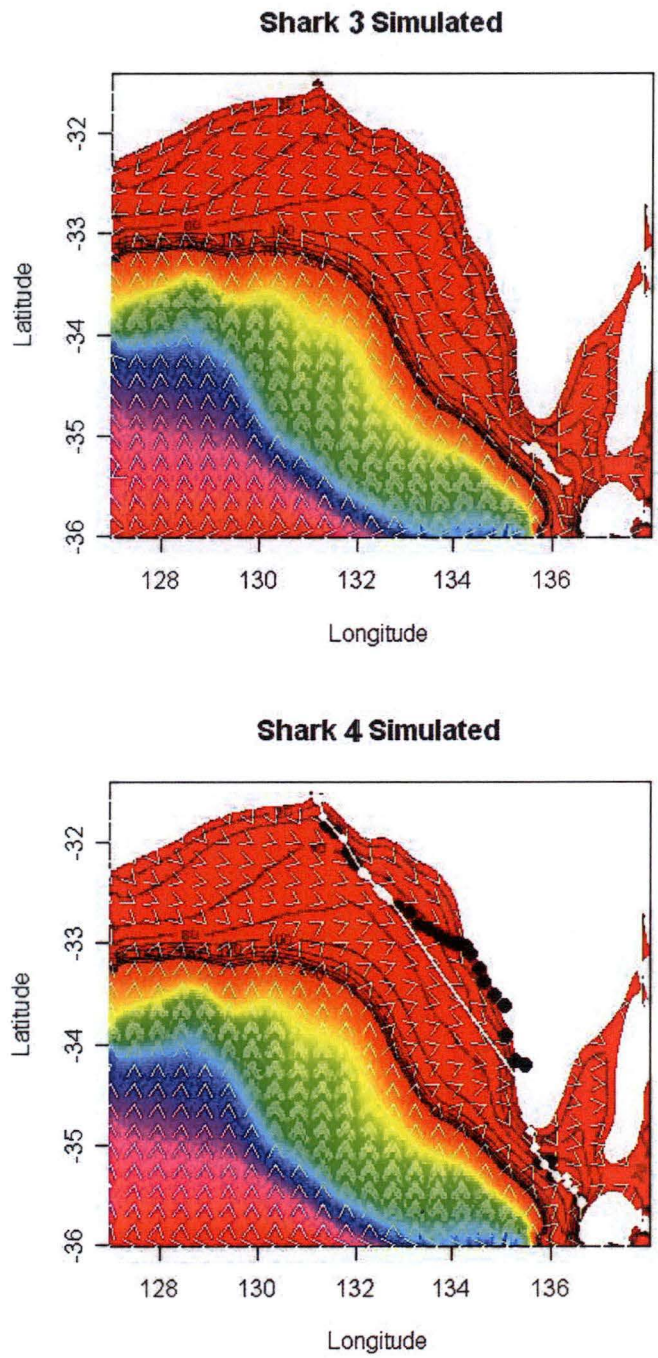
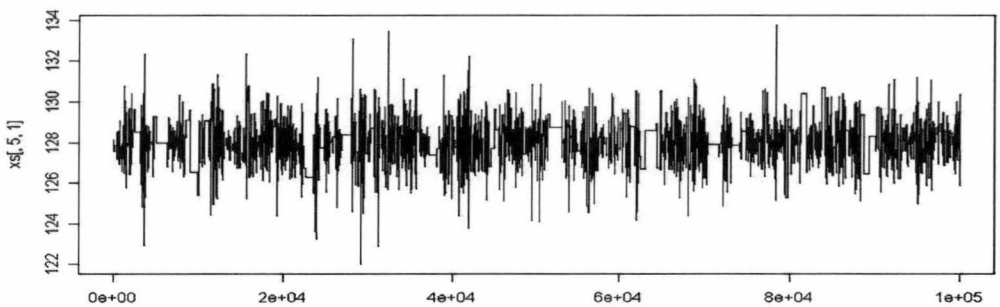
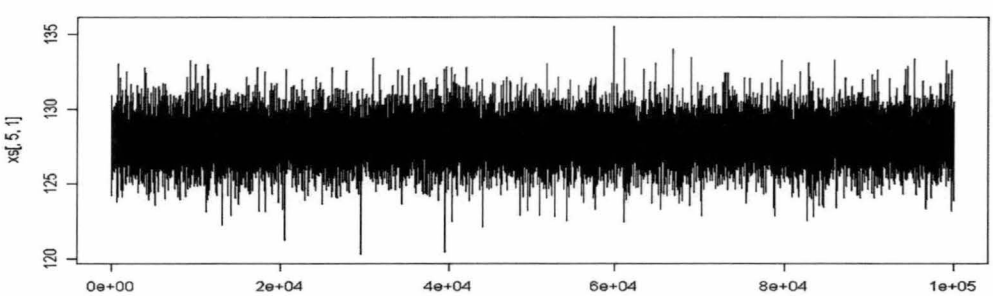


Figure 8.7: *Continued from previous page.* Simulated tracks (shown as connected white dots) that used the observed start location, number of locations and time steps observed for four white sharks (identified by name and number) tagged off South Australia. The simulated path is shown as black dots, the 'observed' locations are white dots joined by a white line. The tracks were simulated using the 'Depth' advection field (shown as white arrows). Bathymetry is indicated by shading and by contours, which are shown every 20m up to a depth of 200m.



(a)



(b)

Figure 8.8: Trace plots (the x-axis shows the MCMC iteration number) for the longitude of the fifth location x_{51} in the path of the observed track for shark 1, Sam (a) before thinning, and (b) after thinning. Results are presented for the model that uses the 'Depth' advection field. The selection of this particular component of the x matrix was arbitrary.

would be required to achieve convergence, requiring prohibitive amounts of computer time. This illustrates the importance of selecting advection fields that do not vary sharply in the region surrounding the track.

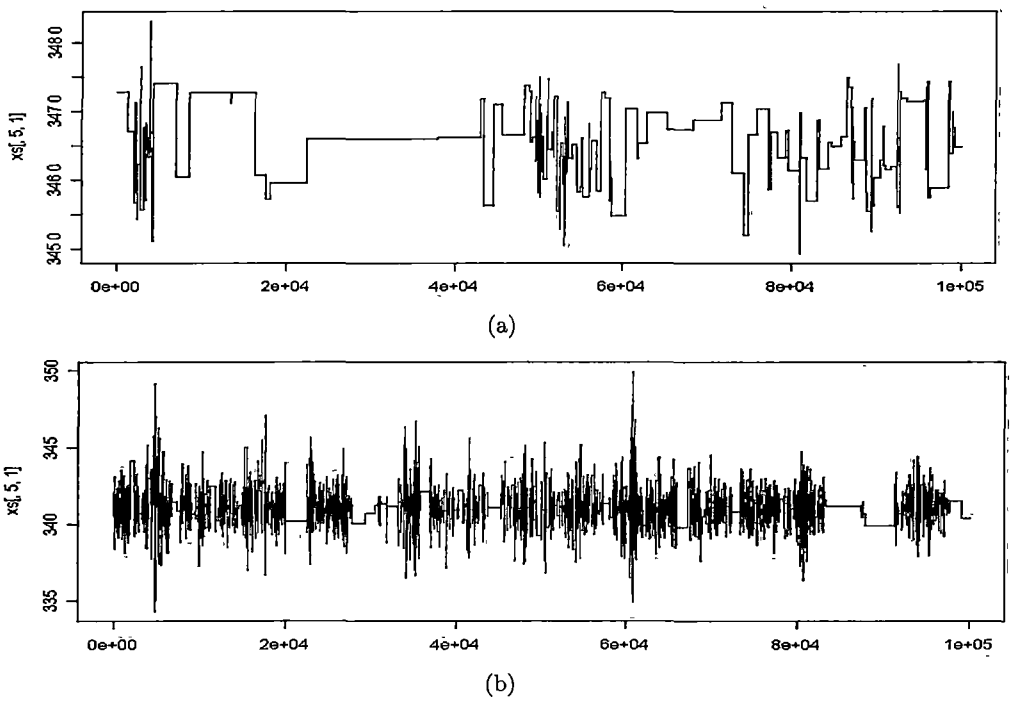


Figure 8.9: Trace plots (the x-axis shows the MCMC iteration number) for the longitude of the fifth location x_{51} in the path of (a) the simulated track, and (b) the observed track for shark 4, Michael. Results are presented for the model that uses the 'Depth' advection field. The selection of this particular component of the x matrix was arbitrary.

8.4.3 Estimation using simulated tracks

In all cases, the DIC selected the correct model ('Depth') when applied to the simulated tracks (Table 8.3). However, the model that uses all three advection fields is a very close second, differing by only 1, 2, and 5 units. Spiegelhalter *et al.* (2002) suggest that models within 1-2 units of the best model deserve consideration whereas those within 3-7 units have considerably less support. In addition, the standard deviations for the DIC statistic for the 'Depth' and 'All' models range between 1.3 and 4.6, so that the difference between their DIC values falls within two standard deviations. Therefore, if this had been a real example in which the true model were unknown, it would not have been possible to discount the 'All' model.

Table 8.3: The mean and standard deviation (in parentheses) of five repeated calculations of the deviance information criterion DIC for the alternative hypotheses, using *simulated* tracks. The 'All' hypothesis combined the three advection fields 'Depth', 'Distance' and 'Direction'. An asterisk indicates the lowest DIC value in each column.

Hypothesis	DIC		
	Track 1	Track 2	Track 3
Depth	650* (2.1)	312* (2.0)	278* (3.0)
Distance	658 (1.2)	392 (0.3)	344 (0.2)
Direction	742 (0.4)	424 (0.1)	344 (0.2)
All	650 (1.3)	313 (1.5)	283 (4.6)

The 'All' model includes the 'Depth' advection field as well as two others, which were not used to generate the simulated tracks. For all simulated tracks, the 95% credibility interval for the posterior for the coefficient of the 'Depth' advection field β_1 in the 'All' model does not include zero, whereas those for the other two fields (β_2 and β_3) do include zero and have their median close to zero (Figure 8.10, righthand-most panels). A zero value for a coefficient of advection indicates that that advection field does not influence movement. Thus the results of the three applications of the 'All' model indicate that the 'Depth' advection field has the greatest influence on movement, however the DIC is unable to separate this model from the one that uses 'Depth' alone, perhaps indicating a lack of parsimony in the DIC statistic in this context.

Note that the negative values estimated for the coefficients β indicate that the shark swims against the advection field. Negative estimates occur for the model that uses all advection fields. As the advection fields are quite similar, negative values for some fields will be balanced by positive values for other fields. In some cases, this will serve to exaggerate the slight differences between the fields, possibly providing a better fit to errors (noise) in the track.

For all simulated tracks the 'All' model suggests that one of β_2 or β_3 has a negative median posterior value, never more negative than -0.25 (Figure 8.10, righthand-most panels). This is balanced by a corresponding slight overestimate of the coefficient β_1. This presumably results from confounding with the inertial term, as discussed in Section 3.5.

The 95% credibility intervals for the posteriors estimated using the correct advection

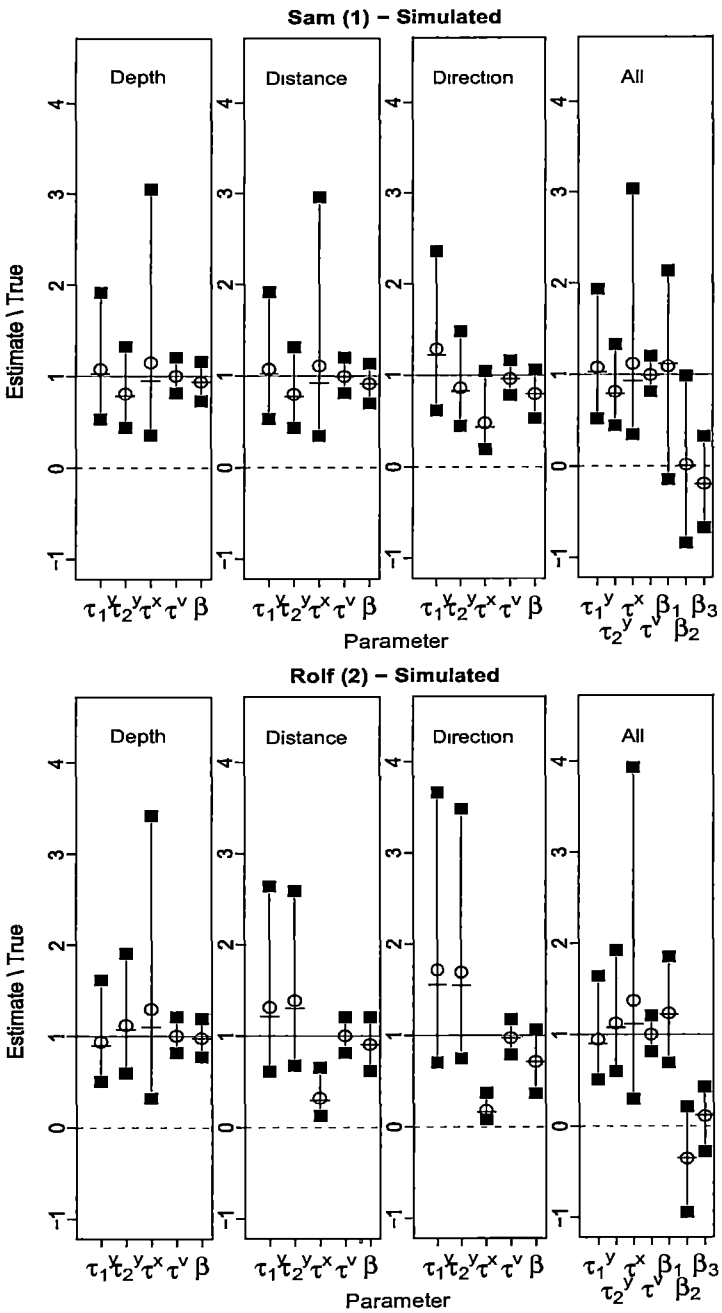


Figure 8.10: Credibility intervals (95%) for parameters estimated from simulated tracks. The 5% and 95%-iles are marked by squares, the median by a dash, and the mean by an open circle. The estimated values are divided by their true (simulation) value so that the grey horizontal line at 1 indicates a correct estimate. The true value for β_2 and β_3 is zero (marked by a dotted line), so that their estimates were divided by the true value for β_1.

Continued overleaf

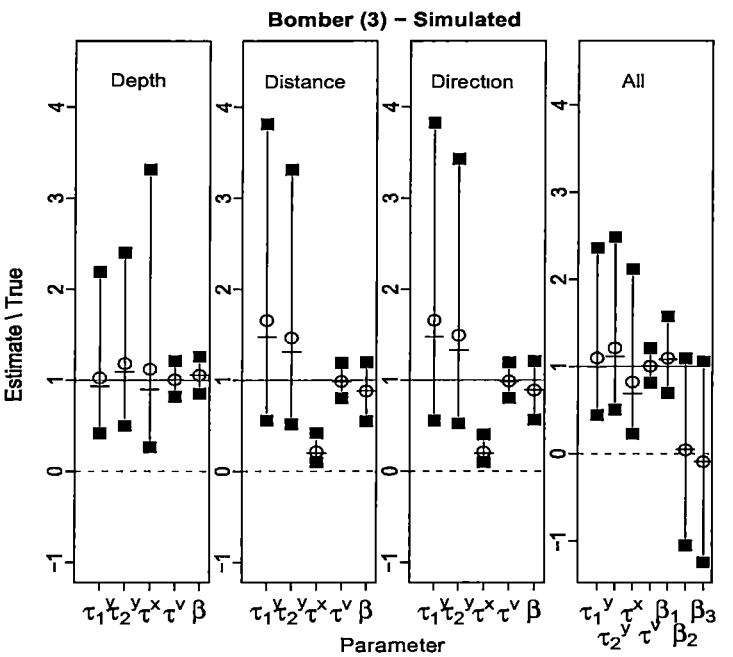


Figure 8.10. *Continued from previous page.* Credibility intervals (95%) for parameters estimated from simulated tracks. The 5% and 95%-iles are marked by squares, the median by a dash, and the mean by an open circle. The estimated values are divided by their true (simulation) value so that the grey horizontal line at 1 indicates a correct estimate. The true value for β_2 and β_3 is zero (marked by a dotted line), so that their estimates were divided by the true value for β_1.

field ('Depth') are largely unbiased (Figure 8.10). When incorrect advection fields are used, the posteriors for the precision parameters are biased when compared with the 'true' values. In particular, that for the track τ^x is biased downwards in most cases, e.g. the two inner panels for sharks 2 and 3. Note that the τ parameters are precisions, not variances, so that these smaller values indicate more variation in the path, allowing greater variability in speed and direction and presumably making up for deficiencies in the ability of these incorrect advection fields to explain the simulated track.

8.4.4 Estimation using observed tracks

When applied to the observed tracks, the hypothesis that indicates travel in a fixed compass direction (the simplest and least explanatory hypothesis) is selected in two cases (Table 8.3) and in the third case (shark 2) it differs by only 1 unit from the 'best' model - the model that uses all advection fields.

Table 8.4: The mean and standard deviation (in parentheses) of five repeated calculations of the deviance information criterion DIC for the alternative hypotheses, using *observed* tracks. The 'All' hypothesis combined the three advection fields 'Depth', 'Distance' and 'Direction'. An asterisk indicates the lowest DIC value in each column.

Hypothesis	DIC		
	Track 1	Track 2	Track 3
Depth	1158 (0.8)	565 (0.4)	553 (0.4)
Distance	1164 (0.2)	562 (0.4)	550 (0.3)
Direction	1150* (0.3)	550 (0.2)	547* (0.5)
All	1152 (0.3)	549* (0.2)	547 (0.5)

For all observed tracks, the difference between the DIC values for the 'All' and 'Direction' models does not deserve consideration, being only 2 units or less. The 'All' model can be ignored - the simulations showed that this model can mimic single hypothesis models by, effectively, eliminating advection fields by setting their coefficient values to zero. Examination of the advection coefficients for the 'All' model (right-handmost plot, Figure 8.11) shows that the 'All' model is, indeed, mimicking the 'Direction' model in this way. For sharks 1 and 2 only the 'Direction' coefficient does not include zero and for shark 3 all coefficients include zero in their 95% CI but this is most peripheral for the 'Direction' CI (β_3). The simulations showed that DIC is insufficiently parsimonious to distinguish two such models.

For track 3, the DIC value for the coastline following hypothesis falls within 3 units of the best model, but otherwise no model that uses a single advection field comes within 7 units of the compass direction model (Table 8.4). The 'Direction' hypothesis is thus given greatest weight by the observed tracks.

The 95% credibility intervals (CI) for the posterior marginals for the simulated tracks were presented as a ratio of their true values (Figure 8.10). The CIs for the observed tracks must be presented without such normalisation, making this plot (Figure 8.11) more difficult to interpret. For greater clarity, the posterior means for the observed tracks are presented in Table 8.5.

Table 8.5: Posterior mean parameter values for four movement hypotheses for each of three observed shark tracks.

Track	Hypothesis	(long)	(lat)	Parameter				
		τ_1^y	τ_2^y	τ^x	τ^v	β_1	β_2	β_3
Shark 1	Depth	1.71	2.07	0.005	9.98	11.22		
Shark 1	Distance	1.78	2.08	0.005	9.97		9.08	
Shark 1	Direction	1.70	2.10	0.006	9.96			10.18
Shark 1	All	1.72	2.08	0.005	9.98	-0.83	4.05	8.24
Shark 2	Depth	2.57	2.76	0.015	9.99	1.77		
Shark 2	Distance	2.55	2.82	0.018	9.98		7.78	
Shark 2	Direction	2.61	2.78	0.015	9.98			3.11
Shark 2	All	2.56	2.81	0.019	9.98	-6.79	3.62	9.31
Shark 3	Depth	1.50	1.65	0.0030	9.96	8.14		
Shark 3	Distance	1.49	1.68	0.0034	9.95		12.15	
Shark 3	Direction	1.45	1.67	0.0033	9.97			10.91
Shark 3	All	1.50	1.64	0.0035	9.96	-4.15	4.88	11.25

Note that the normalisation of the results shown in Figure 8.10 can lead to confusion when compared with the unnormalized results in Figure 8.11. For example, many of the advection coefficient β posteriors have a mean value of roughly 10 for both the simulated (see the true value in Table 8.1) and observed posteriors, but these appear to have a value of 1 in Figure 8.10 and a value of 10 in Figure 8.11. Similarly, the precision for velocity τ^v, which has also has a mean of roughly 10 for both simulated and observed results, appears to be only 1 in Figure 8.10 because the estimated value has been divided by the true value (of 10, Table 8.1).

All observed tracks show strikingly small posterior values for the precision of the track τ^x (Table 8.5), indicating greater variation in the observed path than that assumed in the simulation. Previous simulations (Section 3.5) have shown that the

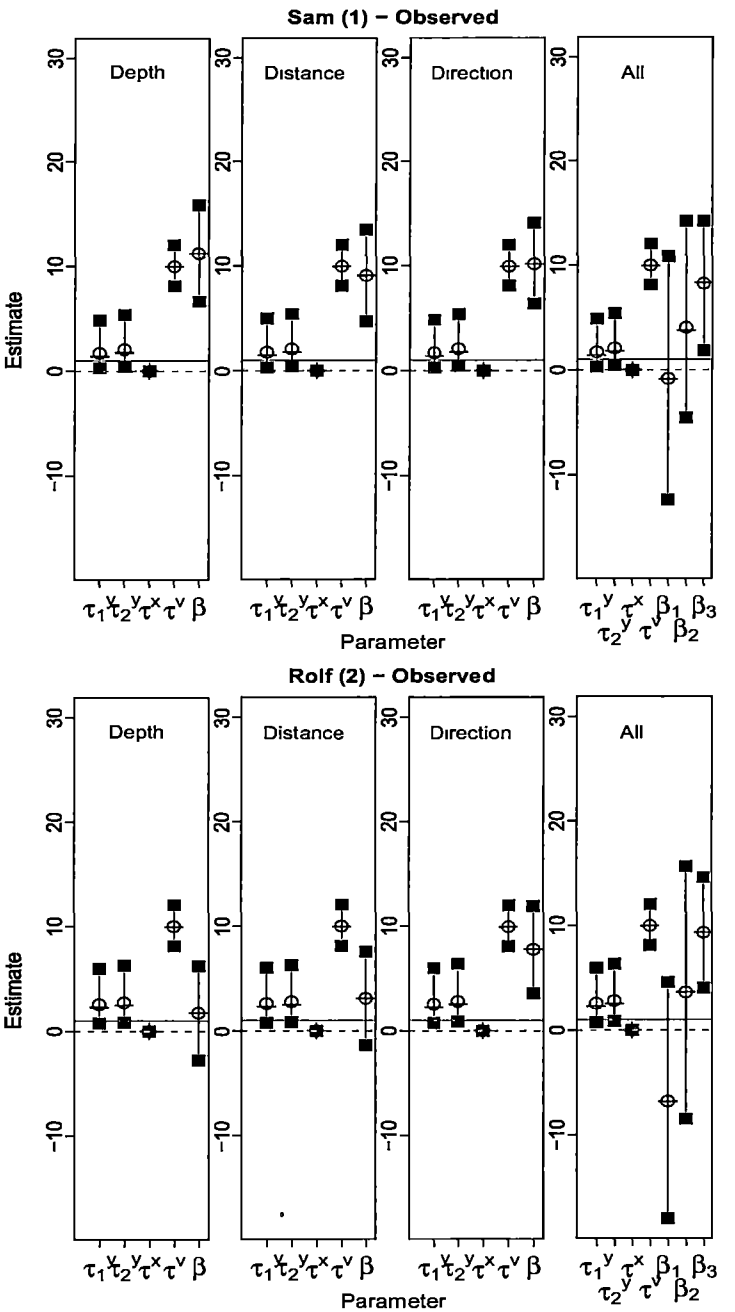


Figure 8.11: Credibility intervals (95%) for parameters estimated from observed tracks. The 5% and 95%-iles are marked by squares, the median by a dash, and the mean by an open circle. A grey horizontal line marks 1 and a dotted line marks zero.

Continued overleaf

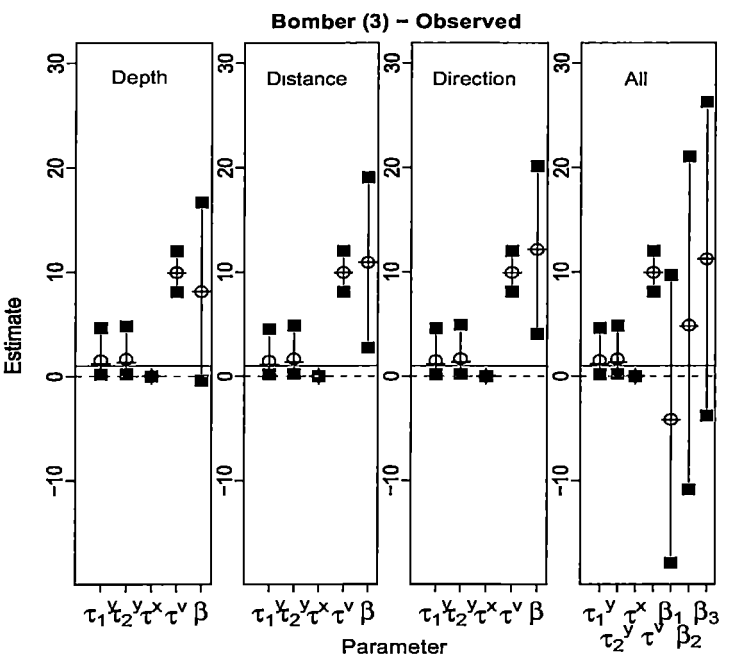


Figure 8.11: *Continued from previous page.* Credibility intervals (95%) for parameters estimated from observed tracks. The 5% and 95%-iles are marked by squares, the median by a dash, and the mean by an open circle. A grey horizontal line marks 1 and a dotted line marks zero.

posterior median for this precision τ^x is, typically, the same as its prior. However, the simulations in this chapter show that when the advection field differs from that used to simulate the track, small values of τ^x are estimated (Figure 8.10).

Similarly, past simulations showed that the mean estimated precision for the observed locations τ^y tends to be similar to its prior mean. However, the results for the observed tracks show greater estimated mean values for the τ^y parameters (ranging between 1.45 and 2) than their posterior mean of 1. This pattern, too, was seen in the simulation results in this chapter (Figure 8.10) for the cases when an incorrect advection field was applied. Further discussion is given in Section 8.5.

The mean posterior values for the precision for velocity τ^v are similar to the prior mean (= 10), even though the prior was no more informative than those used for the other precision parameters (Table 8.2).

8.5 Discussion

The calculations presented in this chapter lead to a number of useful lessons regarding the use of this modelling framework, and pointed the way to future work. One such lesson is that the DIC should be used to compare models that each implement a single hypothesis regarding movement (although each hypothesis might use more than one advection field to achieve a combined effect). Models that incorporate competing hypothesis should not be included in the DIC comparison. The results of such models might closely resemble those of single-hypothesis models (due to advection coefficient estimates of zero) and DIC appears not to be sufficiently parsimonious to distinguish such models.

When comparing models that each implemented a single movement hypothesis, the DIC statistic proved a successful model comparison tool for this movement modelling framework. In simulation, the DIC value for the correct model ('Depth') was smaller by more than 2 standard deviations from the next smallest model.

The proposal distribution used to sample for the path x using the Metropolis-Hastings algorithm is motivated by the assumption that advection is locally constant in the region of each location. When this assumption is violated, as it was by the assumption that advection over land has a southerly flow whereas a more easterly or westerly flow was used in coastal regions, then this Metropolis-Hastings step be-

comes inefficient. This resulted in the need for some thinning of the MCMC chain for shark 1, Sam. For shark 4, which approached land even more closely, having some simulated locations on land, the amount of thinning required was prohibitive. This could have been avoided by not introducing the sharp change in the direction of advection over land. Alternatively, rejection sampling, within the MCMC routine, could be used to prevent the selection of locations that are on land, as used by Sumner *et al.* (n.d.). An advection field based on Laplaces's equation can guide modelled animals around obstacles such as land, as shown in Chapter 7. This methodology was not used in this initial investigation so as not to complicate the behavioural hypotheses with concerns regarding choices of boundary condition. However, this is another avenue for future investigation.

Likewise, abrupt changes in advection fields in the areas where sharks do move should be avoided. In this investigation an abrupt change in the north-south direction across the 80m isobath, and the 50m distance from shore, were avoided by allowing an easterly or westerly flow to dominate. Discontinuities in the bathymetry data, which could also lead to sudden direction changes in an advection field based on bathymetric slope, were smoothed away.

Although abrupt changes in advection fields can be avoided, it would be desirable to have a proposal distribution that is more robust to such discontinuities. Some avenues for future investigations into such proposal distributions are discussed in Section 9.1. However, these are likely to be less efficient than the proposal distribution used in this chapter when advection fields are locally constant, so that the selection of smoothly varying advection fields, where this is possible without loss of realism, will continue to be preferable.

Assumptions had to be made when converting bathymetry to an advection field pertaining to a preferred isobath. First, a particular isobath had to be chosen (80m in this case, Barry Bruce *pers comm.*). This is discussed in the next paragraph. Second, an advection field that merely directs the animal towards this isobath is not sufficient - an overall direction of movement had to be imposed. This could have been done by using two advection fields, one representing movement towards the 80m isobath, and another representing movement due east or due west. The model would estimate coefficients for each field, and would therefore have estimated the relative weighting of these fields. However, the use of two advection fields to represent a single movement hypothesis would have complicated the assessment of competing hypotheses by combining them all into a single model. In this chapter, the isobath-directed and east-west motion were combined into a single field, with

their relative weight specified through the parameter W. This facilitated inference using the 'All' model. However, it would be advantageous to perform subsequent calculations that use two advection fields so that the model can estimate the relative weighting between the two fields. The same is true of the coastline following hypothesis, for which a weight $\widetilde{W}$ was also used (see equation 8.11).

In addition to the W parameter, the co-model that converts bathymetry to an advection field has a parameter that specifies which isobath the animal will follow. The co-model for coastline following, similarly, has a parameter that specifies the animal's preferred distance from shore (50km, in this chapter). Because these co-models were executed prior to the movement model, these parameter values (80m depth and 50km from shore) had to be chosen beforehand and might not have been optimal. Running the co-model alongside the movement model would allow the parameters of the co-model to become estimated parameters. This has not yet been pursued.

The unit in which the time steps Δ_t are measured (e.g. minutes,hours or days) is unimportant in terms of parameter estimation. However, it does affect the scale of other parameters that affect the speed at which the tracked individual moves. These are the coefficients β and the precision parameters for the path τ^x and for velocity τ^v. The coefficients β scale the average speed at which the tracked individual moves as well as dictating how strongly the animal is influenced by advection (an inherent confounding). The precision parameters affect the variability in speed. The time steps scale all of these parameters so that a larger mean time step measure would result in a smaller mean value for the other parameters. This would not affect the path estimates or the model comparison results but should be considered when prior values and starting points for the MCMC routine are chosen, particularly for the precision parameters because their prior mean has been shown to influence their posterior mean.

8.5.1 Application to observed white shark tracks

The application of the movement modelling framework developed in this thesis to observed tracks for white sharks in South Australia shows that this framework, along with the use of the DIC statistic for model comparison, was successful in evaluating the ability of hypothesized movement behaviours to explain observed tracks. In this case, two explanatory navigational behaviours (isobath following and coastline following) were proposed, along with a descriptive hypothesis (movement

in a compass direction). The descriptive hypothesis was most consistent with the observations, indicating that neither of the explanatory hypotheses provided an adequate fit to the data. Simulation showed that the framework was, gratifyingly, able to distinguish between the relatively similar advection fields used in this application.

The modelling framework can be used in two ways to make inference about movement behaviour. First, DIC was successful in selecting the correct model (out of all models that used a single advection field) despite the similarity of the advection fields used by those models. Second, examination of the marginal posteriors for the advection coefficients β for the model that included all three advection fields showed that 95% CIs for the advection fields that were not used to simulate the track included zero, unlike the CI for the advection field that was used to simulate the track.

Application of this modelling framework, and model comparison, to observed shark tracks shows that there is little support for the hypothesis that sharks travel along a preferred isobath, nor for the hypothesis that they remain a preferred distance from the coast. The data was most consistent with a descriptive hypothesis that simply states that sharks move in a southwesterly or northeasterly direction when in Spencer Gulf, and in a northwesterly or southeasterly direction when westwards of Spencer Gulf. For shark 2, which moves north into Spencer Gulf, this was the obvious conclusion because neither of the explanatory advection fields allowed for such movement. For sharks 1 and 3, however, this result could not have been predicted by simply looking at the data and the directions of the advection fields.

The low estimates for τ^x, and the deviation in the estimates of τ^y (between 1.5 and 2.8) from their prior mean (of 1), are at odds with previous simulations (Chapter 3) which have shown that the mean prior for these precision parameters usually dictates the size of their estimates. Interestingly, the same pattern is seen in the simulation results, particularly for sharks 2 and 3 (Figure 8.10), but only in the two cases where the wrong advection field is used. A lower precision for the path allows the model to more easily accommodate variation from the predicted behaviour so that a low τ^x may indicate that the advection field in use is not consistent with the track. However, labelling an estimated value as 'low' actually suggests that the value is downwardly biased, which cannot be known without knowing the true value of the parameter. One could consider that, for a precision parameter, a difference between the estimated value and the mean of the prior indicates a bias. However, this is not a conclusive indication of bias because other influences could cause such a difference, for example, the choice of unit for the time step affects the size of the precision estimates through its influence on the speed of the animal.

It is interesting to note that the observation precisions for longitude and latitude are very similar, although the means are slightly smaller for longitude (indicating greater variation), despite the findings of others that measures of longitude are much more variable than measures of latitude (Sibert *et al.* , 2003; Vincent *et al.* , 2002). The reason for the upward bias in the observation precision τ^y that was found, in simulation, to accompany the downward bias in τ^x is unclear. Often, observations of location were followed by further observations made within minutes or seconds of one another. In this investigation only the first observation in any 15 minute period was retained. This was done to prevent implicit weighting of certain locations and to circumvent difficulties that occurred due to shortage of computer memory (a problem to which the R language is prone). These observations could have provided valuable information on observation precision. By treating 15 minutes intervals as single periods of time during which multiple measures of a single location occur, the problem of implicit weighting of locations could be overcome. Improvements to the computer code, by storing less information in the computer memory and more on disk could circumvent any further memory problems.

The indications are that none of the hypotheses presented here is a good explanation of the behaviour of the tracked animals. This may result from suboptimal choices for the preferred isobath (80m) and preferred distance from shore (50km), or it may indicate that the sharks are using a different navigational method from those proposed, or it may be due to the model not taking account of behavioural switching between different navigational strategies (different behavioural modes).

A suboptimal choice for the preferred depth or distance from shore seem unlikely to give sufficient cause for the selection of the least explanatory advection field, 'Direction'. In particular, the distance from shore (50km) was selected after examination of the observed tracks. Nevertheless, this potential problem could be remedied by allowing the parameters of the co-models to be estimated within the movement model so that the values that give a best fit to the track can, in principle, be found.

It is possible that the sharks may be using a different navigational strategy from those proposed here. Navigation may, for example, use the earth's magnetic field, or may use landmarks. Klimley (1993) found an association between the movements of hammerhead sharks and slope changes in the magnetic intensity of magnetic anomalies in the area, rather than topographic features, whereas Klimley *et al.* (2002) found that individuals belonging to three species of sharks moving through a canyon off the coast of California appeared, at times, to be loosely associated with topo-

graphic features. Magnetic anomalies are often associated with geographic features, such as large under-sea mountains of iron-containing rock, so that the two hypotheses could often be confounded.

Behavioural switching seems to be evident in the observed tracks. Some animals travelled directly from one place to another, and then remained at that place for some time, then travelled on to another location. When travelling, the sharks move rapidly and presumably use some navigational method. When remaining at a location they presumably use a different navigational method to prevent inadvertent straying away from that location. Speed differs between these behavioural modes. Some sharks were seen to reach an eastern-most or western-most point and then turn around and begin retracing their steps. These tracks were truncated at the turnaround point but a behaviour-switching model would have allowed the use of the full tracks. State switching models (such as the Hidden Markov model discussed in Chapter 2 may overestimate the frequency at which switching occurs. A preferable method might be to use mixtures of normal distributions.

Chapter 9

General conclusion

Databases of tracking data are accumulating around the world (Coyne & Godley, 2005) but the quantitative techniques necessary to use these tracks to make inference about animal movement behaviour and its underlying motivations are, as yet, at an early stage of development. This thesis presents a Bayesian state space modelling framework capable of measuring the support given by track data to hypotheses regarding movement behaviour. The framework is innovative because it uses advection fields to facilitate the inclusion of a wide range of hypothesized behaviours; it also caters for unequal time steps, latent location data, and a tendency for directional persistence (inertia). An MCMC algorithm is used for inference so that nonlinear and non-Gaussian models are accommodated. DIC provides a tool for model comparison. Implementation in the R software package (R Development Core Team, 2007) provides flexibility for use of large datasets and coding of co-models that give rise to the advection fields used by the movement model.

State space models (SSM) show great promise in the area of animal movement modelling because of their flexibility and because they can include both observation and process error (Harvey, 1990; Patterson *et al.* , 2007). Both of these are important sources of variation in the context of modelling movement behaviour. Track data includes location errors that may range from hundreds of meters to whole degrees of latitude or longitude (Vincent *et al.* , 2002; Sibert *et al.* , 2003). The process of movement behaviour is, as yet, poorly understood, especially in a context that involves many individuals that can react in different ways.

Movement behaviours, constituting responses to internal states of the individual, or to external environmental cues, are represented in this thesis using advection fields. These are calculated using co-models, which combine hypotheses regarding how an

animal might behave with, where relevant, environmental data. Each co-model, leading to one advection field, constitutes a hypothesis regarding an aspect of movement behaviour. However, movement can be seen as a combined response to several behaviours. A linear combination of advection fields (behaviours) can be created, using weights (coefficients) that are model-estimated parameters. Behaviour can also be allowed to change over time by means of dynamic coefficients so that some behaviours can become more dominant during the tracking period while others fade.

The use of a Bayesian framework results in the estimation of a posterior probability for the location of the animal, rather than a single optimal track. Bayesian estimation also provides a posterior distribution for all other model parameters and states. Such a probabilistic result can facilitate risk assessment, for example, when estimating the potential for overlap between seabird distributions and fishing operations (Broekhuizen *et al.* , 2003).

Using this modelling framework, inference regarding movement behaviour is possible in two ways. First, inference can be achieved through examination of the marginal posterior distributions for the advection coefficients. Greater weight indicates greater influence by a particular advection field (and therefore hypothesized behaviour) on movement behaviour. However, confounding can occur between the coefficients for advection fields that flow in similar directions over the area through which the tracked animal moves, and confounding can also occur between these parameters and inertia, a tendency for directional persistence. Second, state space models that use different (sets of) advection field(s) can be applied to the same observed track and the support given by the data to each alternative model, and therefore hypothesis regarding movement behaviour, assessed. In simulation, the deviance information criterion DIC was able to identify the model that was used to simulate the data. The Bayes factor, a theoretically well-grounded Bayesian model comparison tool, may give support to the wrong model when priors are highly diffuse. ,Bayes factors are also more computationally challenging ,because a difficult integral must be estimated, whereas DIC is easily computed from a set of MCMC draws from the posterior. Informative priors are unlikely to be available in the nascent field of movement behaviour so that DIC is to be preferred to Bayes factors as a model comparison tool in this context. DIC should be used to compare models that each implement a single hypothesis regarding behaviour (implemented as one of more advection fields). Nested models should not be compared using DIC because calculations presented in Chapter 8 show that DIC may not be sufficiently parsimonious in this context, preferring the more complicated model.

Unfortunately, interesting hypotheses exist that cannot be tested using track data because the covariate information is unavailable. In particular, this includes responses to other animals that are not observed or tracked. Individuals are likely to move in response to predators or prey and may respond to others of their own species, either by following or by avoiding them.

It is conceivable that hypotheses will exist that lead to almost identical behaviour, which cannot be discriminated. However, this modelling framework has been successful in discriminating hypotheses regarding movement that lead to very similar behaviours (Chapter 8).

To date, the development of state space models applied to track data has concentrated on inference and on accounting for the error structure of tagging data (Sibert *et al.* , 2003; Jonsen *et al.* , 2005, 2003). These authors have stressed the suitability of the SSM framework for modelling movement behaviour, in particular because it models both process and observation errors. The suitability of a Bayesian approach has also been emphasized because it allows the incorporation of prior information, where available. It simplifies inference for SSMs, avoiding the integration of the likelihood, and does not require that the ratio between the observation and process error variances be specified.

The observation errors typical of satellite tracking data might best be represented using t-distributed errors as well as multiple error categories (Jonsen *et al.* , 2005, 2006). The framework presented in this thesis used normal observation errors to simplify the problem of calculating the conditional distributions used in Gibbs sampling. However, the Metropolis-Hastings within Gibbs algorithm used here could be extended to accommodate the error structure more appropriate to satellite tracking data. This approach to inference has facilitated the use of a nonlinear state equation.

Unusually, the modelling framework presented here accounts for the duration of each time step in the state equation, allowing error to propagate in accordance with the duration of the step and scaling advection and inertia to the duration of the step. This explicitly accounts for the unequal time steps typical of satellite tracking data. SSMs naturally model equally-spaced observations, a problem that other authors have approached by interpolating the data points (Flemming *et al.* , 2006; Jonsen *et al.* , 2006; Tremblay *et al.* , 2006), or by introducing an interpolation into the observation equation Jonsen *et al.* (2005).

Published applications of SSMs to track data have avoided complex hypotheses re-

garding behaviour. SSMs have been used to refine estimates of true location (Sibert *et al.* , 2003; Sumner *et al.* , n.d.). Flemming *et al.* (2006) examined the deviation of turtle's paths from a theoretically ideal great circle route, estimating from this the each animal's degree of navigational confusion. Jonsen *et al.* (2006) and Jonsen *et al.* (2007) showed that foraging and migrating turtles follow different behavioural patterns during the day than at night. Behavioural changes are interpreted through changes, i.e. state-switching, in the variance, or speed parameters of these models. Using this approach to modelling behaviour, it is unclear how arbitrarily complex behaviours might be introduced. The advection field approach presented in this thesis provides a flexible means for incorporating behaviours of almost any level of complexity into an SSM applied to track data.

9.1 Future work

An important addition to the model framework would be the ability to combine data from several tracks in a hierarchical framework. This would allow several animals that are all considered to be moving in response to the same cues to be included in the model. Each would have parameter values drawn from a hyperdistribution so that individuals would be allowed to differ and yet would behave in a broadly similar way. The inclusion of several tracks, in this way, would increase the power of the model to discriminate amongst advective fields (Jonsen *et al.* , 2006).

The Metropolis-Hastings step used for drawing values for the path locations $\boldsymbol{x}$ uses, for the proposal distribution, the distribution that would be the full conditional for $\boldsymbol{x}$ if advection were the same at all possible locations. Provided that advection changes smoothly with location, so that it can be considered to be locally constant, this proposal distribution performs well, as shown in Chapters 3, 4, 5, and 8. When advection is constant, this Metropolis-Hastings step becomes a Gibbs sampling step, never rejecting any proposed set of values, and is therefore highly efficient. It is therefore preferable to use smoothly varying advection fields.

However, the investigator may not wish to expend the additional effort that may be required to ensure a smoothly varying field. For example, it may be hypothesized that the animal's movements are a response to an environmental variable. This variable may have natural discontinuities, like the bathymetry data used in Chapter 8, or observation errors may have introduced noise into the environmental data. It may be necessary to smooth such environmental data prior to calculating a set of

advection fields that represent a movement response to this data. It is also conceivable that behaviours may be hypothesized that cannot, without loss of realism, be rendered smoothly. For these reasons, it is desirable to have, as an alternative to the proposal distribution presented in this thesis, one that is more robust to sharp direction changes in the advection field. One way to do this, albeit at the cost of greater computational time, would be to make draws for each location separately, instead of drawing all at once. Alternatively, the red-black method of drawing every second location, resulting in two blocks, more efficiently exploits the loss of memory in the chain that causes locations 1 and 3 to be less correlated than locations 1 and 2. Even doing this, difficulties are likely to occur for those locations situated in regions where advection changes sharply. A series of diffuse bivariate normal proposal distributions, each centred on an observed location, could be used as proposal distributions. An alternative is to use a modified form of the 'conditional' distribution for x, used in this thesis. For instance, the covariance term calculated for this multivariate normal 'conditional' for x could continue to be used, but the mean could be replaced, either by the observed locations, or by the locations accepted in a previous MCMC step. Yet another alternative is to avoid the MCMC step for x altogether by using one of the modified Kalman filters developed for nonlinear problems to estimate the state variables x and v (Wang, 2007; Carter & Kohn, 1994).

The incorporation of t-distributed errors and error categories, such as those used by Sibert *et al.* (2003) and Jonsen *et al.* (2003) would facilitate application of this modelling framework to Argos tracking data. The computational burden of calculating the required conditional distributions would, however, be much greater.

A version of the movement model was presented in Chapter 4 that allows movement behaviour to change over time through dynamic advection coefficients. These advection coefficients were estimated with difficulty, showing wide credibility intervals and some bias due to confounding with the coefficients for different fields and due to lack of information in the data for coefficients near the start and end of the track. This model could be improved, to some extent, by constraining the coefficients to be positive. This could be achieved through a truncated prior.

In the simulations presented in this thesis, the co-models that calculate the advection fields were implemented prior to implementing the main model so that the advection fields were static, and could only be weighted by the movement model. It is possible, instead, to incorporate the co-model into the main model so that the parameters of the co-model can be estimated together with those of the main model. In this way, the data could be used to inform the advection fields. This

would prevent the rejection, by model comparison, of hypothesized behaviours that are broadly correct but that have been implemented using non-optimal parameters. For example, an animal may navigate by following a preferred line of bathymetry. If it follows the 80m isobath but a co-model is implemented using the 100m isobath, then the resulting advection field may not adequately describe the behaviour of the animal, even though it is correct in essence.

Application of this model to observed track data for white sharks revealed a tendency for sharks to switch between behavioural modes. Sharks moved in a highly directed manner between locations (presumably foraging areas) and then remained at these locations for some time (presumably feeding). Speed of travel differed greatly between these 'travelling' and 'resident' behavioural modes and it is likely that navigational behaviour did too. in future, mixtures of normal distributions could be used to simulate different behavioural modes.

A single investigation presented in Chapter 3 suggested that an asymmetric posterior for velocity might give an indication of incorrect specification of advection fields. Such posteriors resulted from an application of the movement model to white shark tracks (Chapter 8). More extensive simulation testing of the effect of incorrect selection of advection fields, and of other aspects of model structure, than that shown in Chapter 3 would reveal whether or not asymmetry in the posterior for velocity can be used as an accurate indicator of an incorrect advection field and, perhaps, could provide information on how that field could be improved.

Further experience with application of this framework to a variety of real situations and hypothesized behaviours would further elucidate its strengths and weaknesses and, it is hoped, contribute to better understanding of animal movement behaviour and to better use of existing track datasets.

Chapter 10

Description, and dimensions, of symbols used in this thesis

Table 10.1: Description of symbols used in this thesis, together with their dimensions.

Symbol	Description	Dimension
Dimensions		
i	Index for Cartesian coordinates of location	$i = 1, 2$
k	Index for time steps	$k = 1 \ldots N$
N	Number of locations	scalar
n	Number of observed locations	scalar
m	Number of advection fields	scalar
States		
x	Path (true locations)	$N \times 2$
x_1	True location at beginning of first time step	1×2
x_k	True location at beginning of time step k	1×2
x_{ki}	ith coordinate of true location at beginning of time step k	scalar
v	Velocity	$(N-1) \times 2$
v_1	Velocity during first time step	1×2
v_k	Velocity during time step k	1×2
v_{ki}	ith coordinate of velocity at beginning of time step k	scalar
B	Coefficients of advection for all time steps	$(N-1) \times m$
B_1	Coefficients of advection for first time step	$1 \times m$

Continued on Next Page...

Table 10.1 continued

Symbol	Description	Dimension
B_k	Coefficients of advection for time step k	$1 \times m$
vec(B)	Stacked columns of B matrix	$m(N-1) \times 1$
Parameters		
β	Coefficients of advection (time invariant)	$m \times 1$
τ^y	Precision for observed locations y	2×2
τ_i^y,τ_1^y,τ_2^y	τ^y for one Cartesian coordinate	scalar
τ^x	Precision for true locations (path) x	2×2
τ_i^x, τ_1^x, τ_2^x	τ^x for one Cartesian coordinate	scalar
τ^v	Precision for velocity v	2×2
τ_i^v,τ_1^v,τ_2^v	τ^v for one Cartesian coordinate	scalar
τ^B	Covariance for B	$m \times m$
θ	Vector of all model parameters	NA
Errors		
ϵ	Errors in observed locations y	$n \times 2$
ϵ_k	Errors in y at time step k	1×2
ϵ_i	ith coordinate for all errors in y	$n \times 1$
ξ	Errors in true locations (path) x	$(N-1) \times 2$
ξ_k	Errors in x at time step k	1×2
ξ_i	ith coordinate for all errors in x	$(N-1) \times 1$
ζ	Errors in velocity	$(N-2) \times 2$
ζ_k	Errors in v at time step k	1×2
ζ_i	ith coordinate for all errors in v	$(N-2) \times 1$
η	Errors in advection coefficients B	$(N-2) \times m$
η_k	Errors in B at time step k	$1 \times m$
Data		
y	Track (observed locations)	$n \times 2$
y_1	Track at beginning of first time step	1×2
y_k	Track at beginning of time step k	1×2
y_{ki}	ith coordinate of track at beginning of time step k	scalar
t_k	Time when observation y_k was made	scalar
Δ_k	Time difference $t_{k+1} - t_k$ (duration of time step k)	scalar
Δ	Time step durations	$(N-1) \times 1$
Λ_x	Diagonal matrix of time steps	$(N-1) \times (N-1)$
Λ_v	Diagonal matrix of time steps, except the last	$(N-2) \times (N-2)$

Continued on Next Page...

Table 10.1 continued

Symbol	Description	Dimension
V	All covariate information (independant of x)	NA
X, $\mathcal{X}$, $\tilde{\mathcal{X}}$	Advection forces: the subset of V operating at x	see below
X_k	Advection forces, $k = 1 \ldots m$	$2 \times m$
X_{ki}, X_{k1}, X_{k2}	Advection forces on ith coordinate during kth time step	$1 \times m$
$\mathcal{X}_i$, $\mathcal{X}_1$, $\mathcal{X}_2$	Advection forces on ith coordinate, kth row of $\mathcal{X}_i$ is ith row of X_k	$(N-1) \times m$
$\tilde{\mathcal{X}}_i$	m advection forces for $N-1$ time steps, m diagonal blocks of the m columns of $\mathcal{X}_i$	$(N-1) \times m * (N-1)$
Operational matrices		
D_x	Differencing matrix for observations	$(N-1) \times N$
D_v	Differencing matrix for velocities and B	$(N-2) \times (N-1)$
D_B	Differencing matrix for vec(B)	$m(N-2) \times m(N-1)$
P	Selects observed locations	$n \times N$
O	Index of observed locations	$n \times 1$
$\mathcal{I}^{x0}$	Expansion vector $= (1,0,0,...,0,0)^T$	$N \times 1$
$\mathcal{I}^{v0}$	Expansion vector $= (1,0,0,...,0)^T$	$(N-1) \times 1$
Q_B	Expansion matrix $= \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix}$	$m(N-1) \times m$
Priors		
x_0	Mean initial location	1×2
τ^{x0}	Precision for initial location	2×2
v_0	Mean initial velocity	1×2
τ^{v0}	Precision for initial velocity	2×2
β_0	Mean of multivariate normal prior for β	$m \times 1$
Σ_0	Variance for prior for β	$m \times m$
a^y	Shape for gamma prior for τ^y	2×1
a^x	Shape for gamma prior for τ^x	2×1
a^v	Shape for gamma prior for τ^v	2×1
b^y	Rate for gamma prior for τ^y	2×1
b^x	Rate for gamma prior for τ^x	2×1
b^v	Rate for gamma prior for τ^v	2×1
B_0	Mean of multivariate normal prior for B_1	$1 \times m$
τ^{B0}	Precision for prior for B_1	$m \times m$
n_0	Degrees of freedom for Wishart prior for τ^B	scalar

Continued on Next Page...

Table 10.1 continued

Symbol	Description	Dimension
S_0	Scale matrix for Wishart prior for τ^B	$m \times m$
Conditional distribution parameters		
μ^β	Mean of conditional for β	$m \times 1$
Σ^β	Variance of conditional for β	$m \times m$
μ^B	Mean of conditional for B	$m(N-1) \times 1$
Σ^B	Variance of conditional for B	$m(N-1) \times m(N-1)$
μ_i^v	Mean of conditional for v_i	$(N-1) \times 1$
Σ_i^v	Variance of conditional for v_i	$(N-1) \times (N-1)$
μ_i^x	Mean of conditional for x_i	$N \times 1$
Σ_i^x	Variance of conditional for x_i	$N \times N$

CHAPTER 11

Posterior for calculation of conditional distributions for Chapter 4

This appendix shows the unnormalized posterior distribution for the model described in Chapter 4, and describes how this can be used to derive conditional distributions for the parameters and states of that model. Calculations for a gamma, a multivariate normal, and a Wishart conditional are given in appendices 12 to 14.

The conditionals presented in Chapter 3 can be derived using the same methodology. The calculations underlying Chapter 4 are described, rather than those of Chapter 3, because the advection coefficients used in Chapter 4 vary with time, making these less straightforward to work with.

11.1 Conditional distributions

The conditional distribution $p(a \mid b)$ for some variable a, given another variable or set of variables b, is given by their joint distribution $p(a, b)$ divided by the distribution for b, $p(b)$, (Sorensen & Gianola, 2002)

$$p(a \mid b) = \frac{p(a, b)}{p(b)}. \tag{11.1}$$

If a represents one of the parameters or states of the movement model given in Chapter 4, and b represents all of the other parameters and states and the data y, then the joint distribution $p(a, b)$ is the posterior.

The posterior is the product of the likelihood and the priors, divided by an unknown normalising constant, the marginal likelihood $m(y)$. The normalising constant for the conditional, $p(b)$, is also unknown. However, the conditional must be a density function, and because priors have been used that are conjugate with the likelihood, the form of this density is known (Diaconis & Ylvisaker, 1979). All one need do is identify all of the terms in the unnormalized posterior that involve the variable a and express these in the required distributional form. It will then be clear, from the form of that distribution, what normalising constant is required. This constant will represent $m(y)/p(b)$.

11.2 Posterior distribution

First, the unnormalized posterior must be expressed in full. In this, and subsequent appendices, dependence on the control data V is suppressed, for clarity, but note that the conditional given for x is not a true conditional when x is a function of X, that subset of the advection V that operates at the locations x. The symbol θ is used to represent the model parameters, the prior $p(\theta)$ has been expanded. The posterior $p(x, v, B, \theta \mid y)$ is proportional to

$$p(x, v, B, \theta \mid y) \propto p(y \mid x, v, B, \theta)\ p(x \mid v, B, \theta)\ p(v \mid B, \theta)\ p(B \mid \theta) \\ p(\tau^y)\ p(\tau^x)\ p(\tau^v)\ p(\tau^B). \quad (11.2)$$

The symbols used in this appendix are described in Chapter 4 and are listed in Appendix 10.

The likelihood of the data given the full parameter set $p(y \mid x, v, B, \theta)$ is derived from the observation equation (equation 4.5) and the normal error distribution for the observations (equation 4.6). Because the errors are independent between time steps, the overall density is the product of the n normal distributions for each observation (recall that there are N time steps but only n, $n \leq N$ of these are observed). In addition, the two Cartesian components of each observation are independent of one another so that the product consists of $2n$ normal distributions

$$
\begin{aligned}
p(y \mid x, v, B, \theta) &= \prod_{i=1}^{2} \prod_{k \in O} p(y_{ki} \mid x_{ki}, \theta) \\
&= (2\pi)^{-n} \mid \tau^y \mid^n \exp\left[-\frac{1}{2} \sum_{i=1}^{2} \sum_{k \in O} \tau_i^y \epsilon^T \epsilon\right]^2 .
\end{aligned}
\tag{11.3}
$$

Here τ^y is a 2×2 diagonal matrix whereas τ_i^y is scalar. The $n \times 2$ vector of observation errors ϵ is given by

$$
\epsilon = y - P x. \tag{11.4}
$$

The distribution $p(x \mid v, \theta)$, which may be termed the prior for the unobserved state x (Schnute, 1994) is derived from the state equation for x (equation 4.3), and the distribution for the errors in x (equation 4.7), multiplied by the prior for x_1, (the initial condition for x, equation 3.11). These errors are independent between time steps as well as between Cartesian coordinates, therefore the probability density is the product of $2\,(N-1)$ normal distributions

$$
\begin{aligned}
p(x \mid v, B, \theta) &= \prod_{i=1}^{2} p(x_{1i} \mid \theta) \prod_{k=2}^{N} p(x_{ki} \mid x_{k-1i},\, v_{k-1i}, \theta,\, \tilde{\mathcal{X}}_{k-1i}) \\
&= (2\pi)^{-1} \mid \tau^{x0} \mid^{\frac{1}{2}} \exp\left[-\frac{1}{2} \sum_{i=1}^{2} \tau_i^{x0}\, (x_{i1} - x_{i0})^2\right] \\
&\cdot (2\pi)^{-(N-1)} \mid \tau^x \mid^{(N-1)} \exp\left[-\frac{1}{2} \sum_{i=1}^{2} \tau_i^x \left(\xi_i^T \Lambda_x^{-1} \xi_i\right)\right] .
\end{aligned}
\tag{11.5}
$$

Here τ^x is a 2×2 diagonal matrix whereas τ_i^x is scalar. The vector of path location errors ξ is given by

$$
\xi_i = D_x\, x_i - \Lambda_x \left(v_i + \tilde{\mathcal{X}}_i \operatorname{vec}(B)\right). \tag{11.6}
$$

The distribution $p(v \mid B, \theta)$, the prior for the unobserved state v, which gives inertia, is derived from the state equation for v (equation 4.4) and the error distribution for

v (equation 4.8), multiplied by the prior for v_1 (the initial condition for v, equation 3.12). Because these errors are independent between time steps as well as between Cartesian coordinates, the probability density is the product of $2\,(N-2)$ normal distributions

$$\begin{aligned} p(v \mid B, \theta) &= \prod_{i=1}^{2} p(v_{1i} \mid \theta) \prod_{k=2}^{N} p(v_{ki} \mid v_{k-1i},\, \theta) \\ &= (2\pi)^{-1} \mid \tau^{v0} \mid^{\frac{1}{2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{2} \tau_i^{v0} \left(v_{1i} - v_i^0\right)^2\right] \\ &\cdot (2\pi)^{-(N-2)} \mid \tau^{v} \mid^{(N-2)} \exp\left[-\frac{1}{2}\sum_{i=1}^{2} \tau_i^{v} \left(\zeta_i^T \Lambda_v^{-1} \zeta_i\right)\right]. \end{aligned} \tag{11.7}$$

Here τ^v is a 2×2 diagonal matrix whereas τ_i^v is scalar. The vector of errors in velocity ζ is given by

$$\zeta = D_v\, v. \tag{11.8}$$

The distribution $p(B \mid \theta)$, the prior for the unobserved state B, the coefficients of advection, is derived from the state equation for B (equation 4.1) and the error distribution for B (equation 4.2), multiplied by the prior for B_1 (the initial condition of B, equation 4.10). Because these errors are independent between time steps, but not between advection sources, the probability density is the product of $(N-1)$ multivariate normal distributions

$$\begin{aligned} p(B \mid \theta) &= p(B_1 \mid \theta) \prod_{k=2}^{N-1} p(B_k \mid B_{k-1},\, \theta) \\ &= (2\pi)^{-\frac{m}{2}} \mid \tau^{B0} \mid^{\frac{N-2}{2}} \exp\left[-\frac{1}{2}\left(B_1 - B_0\right) \tau^{B0} \left(B_1 - B_0\right)^T\right] \\ &\cdot (2\pi)^{-\frac{m(N-2)}{2}} \mid \Lambda_v^{-1} \otimes \tau^{B} \mid^{\frac{1}{2}} \exp\left[-\frac{1}{2}\eta^T \left(\Lambda_v^{-1} \otimes \tau^{B}\right) \eta\right] \end{aligned} \tag{11.9}$$

where τ^B is an $m \times m$ matrix that may have non-zero off-diagonal elements. The vector of errors in the advection coefficients η is given by

$$\eta = D_v B. \tag{11.10}$$

The gamma prior distributions for the three precision terms τ^y, τ^x, and τ^v are all of the same form. Thus we present this density only once, using τ to represent any one of these three precision terms and a and b to represent the parameters of their prior distributions a^y, a^x, a^v, and b^y, b^x, b^v. The subscript $i = 1, 2$ denotes Cartesian coordinate

$$p(\tau) = \Gamma(a_i)^{-1}\, b_i^{a_i}\, \tau_i^{a_i - 1}\, \exp\left[-b_i\, \tau_i\right]. \tag{11.11}$$

The Wishart prior distribution for the precision for the advection coefficients τ^B is

$$p(\tau^B) = \frac{\left|\tau^B\right|^{\frac{n_0 - m - 1}{2}}\, \exp\left\{-\frac{1}{2}\, \text{trace}\left(S_0^{-1}\, \tau^B\right)\right\}}{2^{\frac{n_0\, m}{2}}\, \left|S_0\right|^{\frac{n_0}{2}}\, \Gamma_m\left(\frac{n_0}{2}\right)}. \tag{11.12}$$

where n_0 and S_0 are the parameters of the Wishart prior (equation 4.9).

Chapter 12

Calculation of gamma conditional distribution for τ_i^x

The precision terms for the observations τ_i^y, the locations τ_i^x, and the velocity τ_i^v used in Chapters 3 and 4 all have gamma priors and corresponding gamma conditional distributions (conjugate with the normal likelihood), (Diaconis & Ylvisaker, 1979). The derivation of the conditionals for all of these precision terms is very similar, therefore only that for τ_i^x is presented.

The full posterior, given in Appendix 11 is the product of $2n$ normal distributions for the observation errors ϵ; $2(N-1)$ normals for the path location errors ξ; $2(N-2)$ normals for the velocity errors ζ; $N-2$ multivariate normals of order m for the errors in the advection coefficients η; 2 normal distributions each for path location x_1 and velocity v_1 during the first time step, and a multivariate normal of order m for the advection coefficients during the first time step B_1; gamma priors for the precisions for observation τ^y, path location τ^x, and velocity τ^v, and a Wishart prior of degree m for the precision for the advection coefficients τ^β.

Terms in τ_i^x appear in two densities. First $p(x \mid v, B, \theta)$, equation 11.5

$$p(x \mid v, B, \theta) \propto (\tau_i^x)^{\frac{N-1}{2}} \exp\left(-\frac{1}{2}\mu_i^T \left[\tau_i^x \Lambda_x^{-1}\right] \mu_i\right) \tag{12.1}$$

where

$$\mu_i = D_x\, x_i \;-\; \Lambda_x \left(\tilde{\mathcal{X}}_i\, \mathrm{vec}(\beta) + v_i\right). \tag{12.2}$$

Note that $\mid \tau^x \mid^{(N-1)}$ in equation 11.5, where τ^x is a diagonal matrix $\mathrm{diag}(\tau_1^x, \tau_2^x)$, is equivalent to the product $\prod_{i=1}^{2} (\tau_i^x)^{\frac{N-1}{2}}$.

Second, is the prior for τ_i^x, ignoring constants

$$p(\tau_i^x) \propto (\tau^x)^{a_i^x - 1}\; \exp\left[-b_i^x\, \tau_i^x\right]. \tag{12.3}$$

Putting these together, and using $\theta(-\tau_i^x)$ to represent all parameters except for τ_i^x

$$p(\tau_i^x \mid \theta(-\tau_i^x), x, v, y, B) \propto (\tau_i^x)^{\frac{N-1}{2} + a_i^x - 1}\; \exp\left[-\frac{1}{2}\tau_i^x\, \left(\mu_i^T\, \Lambda_x^{-1}\, \mu_i - b_i^x\right)\right]. \tag{12.4}$$

The form required for a gamma distribution, again ignoring constants, is

$$p(\tau_i^x \theta(-\tau_i^x), x, v, y, B) \propto (\tau_i^x)^{\hat{a}_i - 1}\; \exp\left[-\hat{b}_i\, \tau_i^x\right]. \tag{12.5}$$

Examining the powers of τ_i^x, it can be seen that

$$\hat{a}_i = \frac{N-1}{2} + a_i^x \tag{12.6}$$

and examining the exponentiated term, it can be seen that

$$\hat{b}_i = -\frac{1}{2}\left(\mu_i^T\, \Lambda_x^{-1}\, \mu_i - b_i^x\right) + b_i^x. \tag{12.7}$$

Chapter 13

Calculation of multivariate normal conditional distribution for B

The path location x and velocity v states and the advection coefficients (β in Chapter 3 and B in Chapter 4) all have normal or multivariate normal conditional distributions (although that for x is a true conditional only if the advection field is constant so that x is independent of advection X). The conditionals for x and v consist of several independent normal distributions, but it is convenient for MCMC sampling to express these as multivariate normals so that the whole state vector may be updated in a single MCMC step. The calculations for deriving these four multivariate normal distributions (for x, v, β and B) are similar, therefore only one is presented. The calculation for B was chosen because it is the most complicated, involving a 'vec' operator, and having the m elements of B correlated within each time step.

Because conjugate multivariate normal priors are used with a normal likelihood, the conditional distribution for B is a multivariate normal (Diaconis & Ylvisaker, 1979). The posterior distribution can be expressed as the product of several distributions. To calculate the conditional it is necessary to isolate those parts of the posterior that pertain to B and show that they can be written in the form of a multivariate normal. There is no need to keep track of terms that are constant with respect to B because once the mean and variance of the conditional distribution have been calculated it will be known, from the form of the multivariate normal density, what the constant for that density has to be.

The full posterior, given in Appendix 11 is the product of $2n$ normal distributions for the observation errors ϵ; $2(N-1)$ normals for the path location errors ξ; $2(N-2)$ normals for the velocity errors ζ; $N-2$ multivariate normals of order m

for the errors in the advection coefficients η; 2 normal distributions each for path location x_1 and velocity v_1 during the first time step, and a multivariate normal of order m for the advection coefficients during the first time step B_1; gamma priors for the precisions for observation τ^y, path location τ^x, and velocity τ^v, and a Wishart prior of degree m for the precision for the advection coefficients τ^β.

In order to update B in a single MCMC step, a multivariate normal conditional for B is required. It is therefore necessary to work, not with the $(N-1) \times m$ matrix B, but instead with the $m * (N-1)$ column vector vec(B), consisting of the stacked columns of B.

Terms that are constant with respect to B can be ignored, leaving terms from only three densities. First, $p(x \mid v, \theta)$, the prior for x (equation 11.5), ignoring terms that are constant with respect to B

$$p(x \mid v, \theta) \propto \exp\left[\sum_{i=1}^{2} -\frac{1}{2}\, \mu_i^T \, \left(\tau_i^x \, \Lambda_x^{-1}\right) \, \mu_i\right] \tag{13.1}$$

where

$$\mu_i = D_x \, x_i \; - \; \Lambda_x \left(v_i + \tilde{\mathcal{X}}_i \, \mathrm{vec}(B)\right).$$

The second and third distributions are components of $p(B \mid \theta)$ (equation 11.9), which is the product of two distributions, both functions of the B. The first of these is the prior distribution describing the initial condition B_1, ignoring constants, and transposing both B_1 and B_0 (which does not affect the scalar result)

$$p(B_1 \mid \theta) \propto \exp\left[-\frac{1}{2}\, \left(B_1^T - B_0^T\right)^T \, \tau^{B0} \, \left(B_1^T - B_0^T\right)\right]. \tag{13.2}$$

However, the terms of this distribution must be expanded to match that of vec(B), and B_1^T must be replaced by a function of vec(B). Therefore the expansion matrix

Q_B is introduced. The product $Q_B\, B_0^T$ is an $m(N-1)$ column vector whose first m elements correspond with B_0^T and whose remaining elements are all zero. The product $Q_B\, \tau^{B0}\, Q_B^T$ is an $m(N-1) \times m(N-1)$ matrix whose top-left $m \times m$ elements correspond with τ^{B0}, and all other elements are zero. Note that the product $Q_B^T\, Q_B$ is the identity matrix of dimension m

$$Q_B^T\, Q_B = I_m. \tag{13.3}$$

Equation 13.2 can be rewritten as

$$\begin{aligned} p(B_1 \mid \theta) &\propto \exp\left[-\frac{1}{2}\left(B_1^T - B_0^T\right)^T\, Q_B^T\, Q_B\, \tau^{B0}\, Q_B^T\, Q_B\, \left(B_1^T - B_0^T\right)\right] \\ &= \exp\left[-\frac{1}{2}\left(Q_B\, B_1^T - Q_B\, B_0^T\right)^T\, Q_B\, \tau^{B0}\, Q_B^T\, \left(Q_B\, B_1^T - Q_B\, B_0^T\right)\right]. \end{aligned} \tag{13.4}$$

Noting that $Q_B\, B_1^T = Q_B\, Q_B^T\, \mathrm{vec}(B)$

$$\begin{aligned} p(B_1 \mid \theta) \propto \exp\Big[-\frac{1}{2}\left(Q_B\, Q_B^T \mathrm{vec}(B) - Q_B\, B_0^T\right)^T\, Q_B\, \tau^{B0}\, Q_B^T \\ \left(Q_B\, Q_B^T \mathrm{vec}(B) - Q_B\, B_0^T\right)\Big]. \end{aligned} \tag{13.5}$$

The second part of $p(B \mid \theta)$, ignoring constants with respect to B, is

$$p(B_k \mid B_{k-1},\, \theta) \propto \exp\left[-\frac{1}{2}\left(D_v\, B\right)^T\, \left(\Lambda_v^{-1} \otimes \tau^B\right)\, \left(D_v\, B\right)\right]. \tag{13.6}$$

This distribution is rewritten by replacing B with $\mathrm{vec}(B)$, and introducing the matrix D_B, which gives the first differences of $\mathrm{vec}(B)$ such that $D_B\, \mathrm{vec}(B) =$

$\text{vec}(D_v\, B)$. This matrix has non-zero elements $(D_B)_{jj} = -1$ and $(D_B)_{j\,j+m} = 1$ for all $j = 1, \ldots, m(N-2)$. Equation 13.6 becomes

$$p(B_k \mid B_{k-1},\, \theta) \propto \exp\left[-\frac{1}{2}\, (D_B\, \text{vec}(B))^T\, \left(\Lambda_v^{-1} \otimes \tau^B\right)\, D_B\, \text{vec}(B)\right]. \quad (13.7)$$

Thus, the posterior distribution consists of a constant with respect to B, multiplied by three exponentiated terms (given by equations 13.1, 13.4 and 13.7). These are all of quadratic form in $\text{vec}(B)$. The multivariate normal conditional for $\text{vec}(B)$ will have the form

$$p(B \mid y, x, v, \theta) \propto \exp\left[\left(\text{vec}(B) - \mu^B\right)^T\, (\Sigma^B)^{-1}\, \left(\text{vec}(B) - \mu^B\right)\right] \quad (13.8)$$

where the mean μ^B and variance Σ^B are to be calculated.

Expanding, and ignoring a constant term

$$p(B \mid y, x, v, \theta) \propto \exp\left[\text{vec}(B)^T\, \Sigma_B^{-1}\, \text{vec}(B) - 2\, \mu_B^T\, \Sigma_B^{-1}\, \text{vec}(B)\right], \quad (13.9)$$

which uses the result $\mu_B^T\, \text{vec}(B) = \text{vec}(B)^T\, \mu_B^T$ because the terms on both sides of the equals sign are scalar and are therefore equal to their transform.

Expanding equations 13.1, 13.4 and 13.7, collecting their quadratic terms, setting these equal to the quadratic term in equation 13.9, and solving for Σ_B^{-1}, reveals that

$$\Sigma^B = \left[Q_B\, \tau^{B0}\, Q_B^T + D_B^T\, \left(\Lambda_v^{-1} \otimes \tau^B\right) D_B + \sum_{i=1}^{2} \tau_i^x\, \tilde{\mathcal{X}}_i^T\, \Lambda_x\, \tilde{\mathcal{X}}_i\right]^{-1}. \quad (13.10)$$

Collecting linear terms in $\text{vec}(B)$ from the expansion of equations 13.1, 13.4 and 13.7, setting these equal to the linear term from equation 13.9, and solving for μ^B reveals that

$$\mu^B = \Sigma^B \left\{ \sum_{i=1}^{2} \left[\tau_i^x \, \tilde{\mathcal{X}}_i^T \, (D_x \, \boldsymbol{x}_i - \Lambda_x v_i) \right] + Q_B \, \tau^{B0} \, B_0^T \right\}.$$

This result is given as equation 4.17 in Chapter 4.

Chapter 14

Calculation of Wishart conditional distribution for τ^B

The precision parameter for the advection coefficients τ^B has a Wishart prior which can be regarded as the multivariate form of the gamma distribution, which was used for the scalar precision terms τ_i^y, τ_i^x, and τ_i^v. The conjugate conditional distribution is also Wishart (Diaconis & Ylvisaker, 1979)

$$p(\tau^B \mid B, \tau^y, \tau^x, \tau^v, y, x, v) \sim \text{Wishart}_m(\hat{n}, \hat{S}) \tag{14.1}$$

where $\hat{n}$ and $\hat{S}$ are quantities that will be calculated in this appendix.

The full posterior, given in Appendix 11 is the product of $2n$ normal distributions for the observation errors ϵ; $2(N-1)$ normals for the path location errors ξ; $2(N-2)$ normals for the velocity errors ζ; $N-2$ multivariate normals of order m for the errors in the advection coefficients η; 2 normal distributions each for path location x_1 and velocity v_1 during the first time step, and a multivariate normal of order m for the advection coefficients during the first time step B_1; gamma priors for the precisions for observation τ^y, path location τ^x, and velocity τ^v, and a Wishart prior of degree m for the precision for the advection coefficients τ^β.

Of these, only two densities that are not constant with respect to τ^B. The first is the Wishart prior for τ^B (equation 11.12), ignoring constants

$$p(\tau^B) \propto \left|\tau^B\right|^{\frac{n_0-m-1}{2}} \exp\left\{-\frac{1}{2}\,\mathrm{trace}\left(S_0^{-1}\,\tau^B\right)\right\}. \tag{14.2}$$

The second is $p(B \mid \theta)$ the multivariate normal prior for B, equation 11.9, ignoring constants

$$\begin{aligned}
&p(B \mid \theta) \\
&\qquad \propto \left|\,\Lambda_v^{-1} \otimes \tau^B\,\right|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\eta^T\left(\Lambda_v^{-1} \otimes \tau^B\right)\eta\right\} \\
&\qquad = \prod_{k=1}^{N-2} \left|\Delta_k^{-1}\tau^B\right|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\eta_k\left[\Delta_k^{-1}\tau^B\right](\eta_k)^T\right\} \\
&\qquad = \left|\prod_{k=1}^{N-2} \Delta_k^{-1}\,\tau^B\right|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\sum_{k=1}^{N-2}\eta_k\,\tau^B\,(\eta_k)^T\,\Delta_k^{-1}\right\} \\
&\qquad = \left|\tau^B\right|^{\frac{N-2}{2}} \prod_{k=1}^{N-2} \Delta_k^{-\frac{m}{2}} \exp\left\{-\frac{1}{2}\,\mathrm{trace}\left[\eta_k\,\tau^B\,(\eta_k)^T\,\Lambda_v^{-1}\right]\right\} \\
&\qquad = \left|\tau^B\right|^{\frac{N-2}{2}} \prod_{k=1}^{N-2} \Delta_k^{-\frac{m}{2}} \exp\left\{-\frac{1}{2}\,\mathrm{trace}\left[(\eta_k)^T\,\Lambda_v^{-1}\,\eta_k\,\tau^B\right]\right\}.
\end{aligned} \tag{14.3}$$

Here $\eta_k = B_{k+1} - B_k$ is a row vector of length m, and $\eta = D_v\,B$ is an $(N-2) \times m$ matrix of first differences of B. The matrix Λ_v is made up of time steps $\Delta_t, t = 1 \ldots N-2$.

Note that the trace could be introduced because the exponentiated term above is a sum of scalar terms, and that the properties of the trace allow the rearrangement of the terms inside.

Putting equations 14.2 and 14.3 together, again ignoring constants with respect to τ^B, and using $\theta(-\tau^B)$ to indicate all parameters except for τ^B

$$\begin{aligned}
p(\tau^B \mid \theta(-\tau^B), x, v, y, B) \propto &\left|\tau^B\right|^{\frac{n_0-m-1}{2}} \exp\left\{-\frac{1}{2}\,\mathrm{trace}\left(S_0^{-1}\,\tau^B\right)\right\} \\
&\left|\tau^B\right|^{\frac{N-2}{2}} \exp\left\{-\frac{1}{2}\,\mathrm{trace}\left[(\epsilon_k)^T\,\Lambda_v^{-1}\,\epsilon_k\,\tau^B\right]\right\}.
\end{aligned} \tag{14.4}$$

The aim, a Wishart density (equation 14.1), again ignoring constants, has the form

$$p(\tau^B \mid \theta(-\tau^B), x, v, y, B) \propto \left|\tau^B\right|^{\frac{\hat{n}-m-1}{2}} \exp\left\{-\frac{1}{2} \text{ trace}\left(\hat{S}^{-1}\,\tau^B\right)\right\}. \qquad (14.5)$$

Examining the powers of τ^B in equations 14.4 and 14.5 it can be seen that

$$\hat{n} = n_0 + N - 2 \qquad (14.6)$$

and from the exponentiated terms it can be seen that

$$\hat{S}^{-1} = S_0^{-1} + (\epsilon)^T\, \Lambda_v^{-1}\, \epsilon \qquad (14.7)$$

as shown in equation 4.16.

CHAPTER 15

R code for Figure 6.1

```
N   <- 1000000      # Number of samples
r   <- 1.5          # RIS variance scaling
mu0 <- 0;           # Prior mean
tau0 <- 0.1         # Inverse variance for prior
set.seed(954)       # Seed for random number generator
ni <- c(1,5,50,500) # Number of data points
loglik <- double(N) # Storage for log-likelihood values
par(mfrow=c(2,2), mar=(c(3,3,2,1)+0.1), oma=c(2,2,1,1)+0.1)

for (i in 1:4) {
  n <- ni[i]
  x <- rnorm(n,0,1)

  # Draw from posterior
  mu <- rnorm(N,(mu0*tau0+sum(x))/(n+tau0),1/sqrt(n+tau0))
  loglik <- colSums(outer(x,mu,dnorm,log=T)) # Likelihood

  ## RIS
  logimp <- dnorm(mu,(mu0*tau0+sum(x))/(n+tau0),r/sqrt(n+tau0)
            ,log=T)-dnorm(mu,mu0,1/sqrt(tau0),log=T)-loglik

  ## Estimated log marginal likelihoods (harmonic mean, RIS, exact)
  logHM  <- min(loglik)-log(cumsum(exp(min(loglik)-loglik))/1:N)
  logRIS <- -log(cumsum(exp(logimp-min(logimp)))/1:N)-min(logimp)
  logM <- (log(tau0/(n+tau0))-n*log(2*pi)-mu0^2*tau0+(mu0
          *tau0+sum(x))^2/(n+tau0)-sum(x^2))/2
```

```
  ## Plot
  xvals <- seq(1/N,1,length.out=N)
  plot(xvals,logHM-logM,type="l",ylim=c(-1,4),col="black"
       ,lty=1,xlab="",ylab="")                              # Harmonic mean
  lines(xvals,logRIS-logM,lty=2); abline(h=0,col="grey")# RIS
  title(paste("n = ",n),cex=1.0)                            # Title
  if (i==1) legend("topright",c("HM","RIS",),lty=c(1,2),col=c(1,1))
}
mtext("Length of MCMC chain (millions)",side=2,outer=TRUE)
mtext("Error in log[m(y)]",side=1,outer=TRUE)
```

REFERENCE LIST

AITKIN, M. 1991. Posterior Bayes factors. *Journal of the Royal Statistical Society B*, **53**(1), 111–142.

AKAIKE, H. 1973. Information theory as an extension of the maximum likelihood principle. *Pages 267–281 of:* PETROV, B.N., & CSAKI, F. (eds), *Second International Symposium on Information Theory.* Budapest: Akademiai Kiado.

ALERSTAM, T. 2003. The lobster navigators. *Nature*, **421**(6918), 27–28.

AMLANER, C.J. JR., SIBLY, R., & MCCLEERY, R. 1979. Effects of telemetry transmitter weight on breeding success in herring gulls. *Pages 254–259 of:* LONG, F.M. (ed), *Proceedings of the Second International Conference on Wildlife Biotelemetry.*

ANDERSON, B.D.O., & MOORE, J.B. 1979. *Optimal Filtering.* Englewood Cliffs: Prentice-Hall.

ANDERSON-SPRECHER, R., & LEDOLTER, J. 1991. State-space analysis of wildlife telemetry data. *Journal of the American Statistical Association*, **86**(415), 596–602.

AOKI, M. 1987. *State Space Modeling of Time Series.* Springer-Verlag.

ATKINSON, R.P.D., RHODES.C.J., MACDONALD, D.W., & ANDERSON, R.M. 2002. Scale-free dynamics in the movement patterns of jackals. *OIKOS*, **98**, 134–140.

AUSTIN, D., BOWEN, W.D., & MCMILLAN, J.I. 2004. Intraspecific variation in movement patterns: modeling individual behaviour in a large marine predator. *OIKOS*, **105**, 15–30.

BARRETT, L., & LOWEN, C.B. 1998. Random walks and the gas model: spacing behaviour of grey-cheeked mangabeys. *Functional Ecology*, **12**, 857–865.

BARTLETT, M.S. 1957. Comment on "A statistical paradox" by D.V. Lindley. *Biometrika*, **44**, 533–534.

BARTSCH, J., REID, D., & COOMBS, S.H. 2004. Simulation of mackerel (*Scomber scombrus*) recruitment with an individual-based model and comparison with field data. *Fisheries Oceanography*, **13**(6), 380–391.

BARTUMEUS, F., DA LUZ, M.G.E., VISHWANATHAN, G.M., & CATALAN, J. 2005. Animal search strategies: a quantitative random-walk analysis. *Ecology*, **8**(11), 3078–3087.

BECKER, E. F. 1991. A terrestrial furbearer estimator based on probability sampling. *The Journal of Wildlife Management*, **55**, 730–737.

BENHAMOU, S. 2004. On the expected net displacement of animals' random movements. *Ecological Modelling*, **171**, 207–208.

BENHAMOU, S., BOVET, P., & POUCET, B. 1995. A model for place navigation in mammals. *Journal of Theoretical Biology*, **173**, 163–178.

BENNETT, D.A., & TANG, W. 2006. Modelling adaptive, spatially aware, and mobile agents: Elk migration in Yellowstone. *International Journal of Geographical Information Science*, **20**, 1039–1066.

BERG, A., MEYER, R., & YU, J. 2004. Deviance information criterion for comparing stochastic volatility models. *Journal of Business & Economic Statistics*, **22**(1), 107–120.

BERGMAN, C.M., SCHAEFER, J.A., & LUTTICH, S.N. 2000. Caribou movement as a correlated random walk. *Oecologia*, **123**, 364–374.

BERTRAND, S., BURGOS J.M. GERLOTTO F., & ATIQUIPA, J. 2005. Levy trajectories of Peruvian purse-seiners as an indicator of the spatial distribution of anchovy (Engraulis ringens). *ICES Journal of Marine Science*, **62**, 477–482.

BEST, N., COWLES, M.K., & VINES, K. 1996. *CODA Convergence Diagnostics and Output Analysis Software for Gibbs Sampling Output. Version 3.0.* Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge.

BEUCHNER, H. K., CRAIGHEAD, F. C. JR., CRAIGHEAD, J. J., & COTE, C. E. 1971. Satellites for research on free-roaming animals. *BioScience*, **21**, 1201–1205.

BEVERTON, R.J.H., & HOLT, S.J. 1957. On the dynamics of exploited fish populations. *Fishery Investigations, Ministery for Agriculture, Fish Food (Great Britain), Series II*, **19**, 533.

BIAN, L. 2003. The representation of the environment in the context of individual-based modeling. *Ecological Modelling*, **159**(2-3), 279–296.

BLOCK, B.A., DEWAR, H., FARWELL, C., & PRINCE, E.D. 1998. A new satellite technology for tracking the movements of Atlantic bluefin tuna. *Proceedings of the National Academy of Sciences of the United States of America (Ecology)*, **95**, 9384–9389.

BOAG, D.A. 1972. Effect of radio packages on behaviour of captive red grouse. *Journal of Wildlife Management*, **36**, 511–518.

BOLES, L.C., & LOHMANN, K.J. 2003. True navigation and magnetic maps in spiny lobsters. *Nature*, **421**(6918), 60–63.

BONFIL, B, MEYER, M, SCHOLL, M.C., JOHNSON, R., O'BRIEN, S., OOSTHUIZEN, H., SWANSON, S., KOTZE, D., & PATERSON, M. 2005. Transoceanic migration, spatial dynamics, and population linkages of white sharks. *Science*, **310**, 100–103.

BORCHERS, D.L., BUCKLAND, S.T., & ZUCCHINI, W. 2002. *Estimating Animal Abundance.* London, U.K.: Springer.

BOUSTANY, A.M., DAVIS, S.F., PYLE, P., ANDERSON, S.D., LE BOEUF, B.J., & BLOCK, B.A. 2002. Expanded niche for white sharks. *Nature*, **415**, 35–36.

BOWLER, D.E., & BENTON, T.G. 2005. Causes and consequences of animal dispersal strategies: relating individual behaviour to spatial dynamics. *Biological Reviews*, **80**, 205–225.

BOX, G.E.P. 1979. *Chap. Robustness in the strategy of scientific model building of:* LAUNER, R.L., & WILKINSON, G.N. (eds), *Robustness in statistics.* Academic Press, New York.

BOX, G.E.P., & JENKINS, G.M. 1976. *Time Series Analysis: Forecasting and Control, Revised Edition.* San Francisco: Holden-Day.

BRASSINGTON, G.B., PUGH, T., C., SPILLMAN, SCHULZ, E., BEGGS, H., SCHILLER, A., & OKE, P.R. 2007. Bluelink development of operational oceanography and servicing in australia. *Journal of Research and Practice in Information Technology*, **39 (2)**, 151–164.

BRILLINGER, D.R., PREISLER, H.K., AGER, A.A., KIE, J.G., & STEWART, B.S. 2002. Employing stochastic differential equations to model wildlife motion. *Bulletin Brazilian Mathematical Society*, **33(3)**, 385–408.

BROEKHUIZEN, N., STAHL, J.C., & SAGAR, P.M. 2003. Simulating the distribution of southern Buller's albatross using an individual-based population model. *Journal of Applied Ecology*, **40(4)**, 678–691.

BROOKS, R.J., & BANKS, E.M. 1971. Radio-tracking study of lemming home range. *Communications in Behavioural Biology*, **4**, 1–5.

BRUCE, B.D., CONDIE, S.A., & SUTTON, C.A. 2001. Larval distribution of blue grenadier (*Macruronus novaezelandiae* Hector) in south-eastern Australia: further evidence for a second spawning area. *Marine & Freshwater Research*, **53**, 603–610.

BRUCE, B.D., STEVENS, J.D., & MALCOLM, H. 2006. Movements and swimming behaviour of white sharks *Carcharodon carcharias* in Australian waters. *Marine Biology*, **150**, 161–172.

BUCKLAND, S.T., NEWMAN, K.B., THOMAS, L., & KOESTERS, N.B. 2004. State-space models for the dynamics of wild animal populations. *Ecological Modelling*, **171**, 157–175.

BURNHAM, K.P., & ANDERSON, D.R. 1998. *Model Selection and Inference: A Practical Information-Theoretic Approach.* Springer-Verlag Telos.

CAIN, S.D., BOLES, L.C., WANG, J.H., & LOHMANN, K.J. 2005. Magnetic orientation and navigation in marine turtles, lobsters, and molluscs: concepts and conundrums. *Integrative and Comparative Biology*, **45**, 539–546.

CAMPBELL, J.B. 2006. *Introduction to Remote Sensing.* New York: Guilford Press.

CAREY, F.G., & SCHAROLD, J.V. 1990. Movements of blue sharks (*Prionace glauca*) in depth and course. *Marine Biology*, **106**, 329–342.

CARLIN, B.P., & CHIB, S. 1995. Bayesian model choice via Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society B*, **57**(3), 473–484.

CARLIN, B.P., & LOUIS, T.A. 2000. *Bayes and Empirical Bayes Methods for Data Analysis.* Chapman and Hall / CRC, Boca Raton, USA.

CARLIN, B.P., POLSON, N.G., & STOFFER, D.S. 1992. A Monte Carlo approach to nonnormal and nonlinear state-space modelling. *Journal of the American Statistical Association*, **87**(418), 493–500.

CARTER, C.K., & KOHN, R. 1994. On Gibbs sampling for state space models. *Biometrika*, **81**(3), 541–553.

CASWELL, H. 2001. *Matrix Population Models.* Sinauer associates, Sunderland, MA.

CELEUX, G., & DIEBOLT, J. 1985. The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Computational Statistics Quarterly*, **2**, 73–82.

CHAKRAVARTI, N. 2004. Beyond Brownian motion: a Levy flight in magic boots. *Resonance*, 50–63.

CHATFIELD, C. 2004. *The Analysis of Time Series: an Introduction. Sixth Edition.* USA: Chapman and Hall / CRC.

CHIB, S. 1995. Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, **90**(432), 1313–1321.

CHIB, S., & GREENBERG, E. 1996. Markov chain Monte Carlo simulation methods in econometrics. *Econometric Theory*, **12**, 409–431.

CHIB, S., & JELIAZKOV, I. 2001. Marginal likelihood from the Metropolis-Hastings output. *Journal of the American Statistical Association*, **96**(453), 270–281.

CLARK, C.W., & ELLISON, W.T. 2000. Calibration and comparison of the acoustic location methods used during the spring migration of the bowhead whale *Balaena mysticetus*, off Point Barrow, Alaska, 1984-1993. *Journal of the Acoustics Society of America*, **107**, 3509–3517.

COLE, B.J. 1995. Fractal time in animal behaviour: the movement activity of *Drosophila*. *Animal Behaviour*, **50**, 1317–1324.

CONGDON, P. 2006. Bayesian model choice based on Monte Carlo estimates of posterior model probabilities. *Computational Statistics and Data Analysis*, **50**, 346–357.

COUGHLIN, D.J., STRICKLER, J.R., & SANDERSON, B. 1992. Swimming and search behaviour in clownfish, *Amphiprion perideraion*, larvae. *Animal Behaviour*, **44**, 427–440.

COWLES, M.K., & CARLIN, B.P. 1996. Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association*, **91**(434), 883–904.

COYNE, M.S., & GODLEY, B.J. 2005. Satellite tracking and analysis tool (STAT) and integrated system for archiving, analysing and mapping animal tracking data. *Marine Ecology Progress Series*, **301**, 1–7.

CRAIGHEAD, J.J., CRAIGHEAD, F.C. JR, VARNEY, J.R., & COTE, C.E. 1971. Satellite monitoring of black bears. *BioScience*, **21**, 1206–1212.

CRITTENDEN, R.N. 1994. A diffusion model for the downstream migration of sockeye salmon smolts. *Ecological Modelling*, **91**, 69–84.

DALGAARD, P., & LARSEN, M. 1990. Fitting numerical solutions of differential equations to experimental data: a case study and some general remarks. *Biometrics*, **46**, 1097–1109.

DE BIE, T, & CRISTIANINI, N. 2004. Kernel methods for exploratory pattern analysis: a demonstration on text data. *Lecture Notes in Computer Science, Springer Berlin / Heidelberg*, 16–29.

DE VALPINE, P. 2002. Review of methods for fitting time-series models with process and observation error and likelihood calculations for nonlinear, non-Gaussian state-space models. *Bulletin of Marine Science*, **70**(2), 455–471.

DE VALPINE, P., & HASTINGS, A. 2002. Fitting population models incorporating process noise and observation error. *Ecological Monographs*, **72**, 57–76.

DEANGELIS, D.L., & GROSS, L.J. 1992. *Individual-Based Models and Approaches in Ecology: Populations, Communities and Ecosystems.* New York: Chapman and Hall.

DELLAPORTAS, P., FORSTER, J.J, & NTZOUFRAS, I. 2002. On Bayesian model and variable selection using MCMC. *Statistics and Computing*, **12**, 2736.

DENNIS, B., PONCIANO, J.M., LELE, S.R., TAPER, M.L., & STAPLES, D.F. 2006. Estimating density dependence, process noise, and observation error. *Ecological Modelling*, **76**(3), 323–341.

DIACONIS, P., & YLVISAKER, D. 1979. Conjugate priors for exponential families. *The Annals of Statistics*, **7**(2), 269–281.

DICICCIO, T.J., KASS, R.E., RAFTERY, A., & WASSERMAN, L. 1997. Computing Bayes factors by combining simulation and asymptotic approximations. *Journal of the American Statistical Association*, **92**(439), 903–915.

DORN, M.W. 2001. Fishing behavior of factory trawlers: a hierarchical model of information processing and decision-making. *ICES Journal of Marine Science*, **58**(1), 238–252.

DURBIN, J., & KOOPMAN, S.J. 2001. *Time Series Analysis by State Space Methods.* Oxford University Press, U.K.

EIANE, K., & PARISI, D. 2001. Towards a robust concept for modelling zooplankton migration. *Sarsia*, **86**, 465–475.

ELSDON, T.S., & GILLANDERS, B.M. 2003. Reconstructing migratory patterns of fish based on environmental influences on otolith chemistry. *Reviews in Fish Biology and Fisheries*, **13 (3)**, 219–235.

EVANS, M. 1988. *Monte Carlo computation of marginal posterior quantiles.* Tech. rept. Department of Statistics, University of Toronto, Canada.

EVANS, M., HASTINGS, N., & PEACOCK, B. 2000. *Statistical Distributions, Third Edition.* USA: Wiley-Interscience.

FARMER, J.D. 2001. Towards agent-based models for investment. *Benchmarks and Attribution Analysis, Association for Investment and Management Research*, 61–70.

FLEMMING, J.E.M., FIELD, C.A., JONSEN, I.D., & MYERS, R.A. 2006. How well can animals navigate? Estimating the circle of confusion from tracking data. *Environmetrics*, **17**, 351–362.

FRANKE, A., CAELLI, T., & HUDSON, R.J. 2004. Analysis of movement and behaviour of caribou (*Rangifer tarandus*) using hidden Markov models. *Ecological Modelling*, 259–270.

FRANKE, A., CAELLI, T., KUZYK, G., & HUDSON, R.J. 2006. Prediction of wolf (*Canis lupus*) kill-sites using hidden Markov models. *Ecological Modelling*, **197**, 237–246.

FREEMAN, S.N., & KIRKWOOD, G.P. 1995. On a structural time series method for estimating stock biomass and recruitment from catch and effort data. *Fisheries Research*, **22**, 77–98.

FRENCH, D.P., REED, M., CALAMBOKIDIS, J., & CUBBAGE, J.C. 1989. A simulation model of seasonal migration and daily movements of the northern fur seal. *Ecological Modelling*, **48**(3-4), 193–219.

FRITZ, H., SAID, S., & WEIMERSKIRCH, H. 2003. Scale-dependent hierarchical adjustments of movement patterns in a long-range foraging seabird. *Proceedings of the Royal Society of London*, **270**(1520), 1143–1148.

FRUHWIRTH-SCHNATTER, S. 2001. Fully Bayesian analysis of switching Gaussian state space models. *Annals of the Institute of Statistical Mathematics*, **53**(1), 31–49.

FRUSHER, S.D., SEMMENS, J.M., HODGSON, K., PECL, G.T., GARDNER, C., HARRINGTON, J., TRACEY, S., & JACKSON, G.D. 2003. Use of acoustic technology to track animal movements in S-E Tasmania. *Fishing Today, Turtle Press, Tasmania, Australia*, **16**, 2.

GAMERMAN, D. 1998. Markov chain Monte Carlo for dynamic generalised linear models. *Biometrika*, **85**(1), 215–227.

GAMERMAN, D., & LOPES, H.F. 2006. *Markov chain Monte Carlo Stochastic Simulation for Bayesian Inference*. Chapman and Hall / CRC, U.S.A.

GAMERMAN, D., & MIGON, H.S. 1993. Dynamic hierarchical models. *Journal of the Royal Statistical Society B*, **55**(3), 629–642.

GAUTESTAD, A.O., & MYSTERUD, I. 2005. Intrinsic scaling complexity in animal dispersion and abundance. *American Naturalist*, **165**(1), 44–55.

GELFAND, A.E. 1996. Markov chain Monte Carlo in practice. Chapman and Hall / CRC.

GELFAND, A.E., & DEY, D.K. 1994. Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society B*, **56**(3), 501–514.

GELFAND, A.E., & SMITH, A.F.M. 1990. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**(410), 398–409.

GELFAND, A.E., HILLS, S.E., RACINE-POON, A., & SMITH, A.F.M. 1990. Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association*, **85**, 972–985.

GELMAN, A. 2006. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, **1**(3), 515–533.

GELMAN, A., & MENG, X.-L. 1998. Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. *Statistical Science*, **13**(2), 163–185.

GELMAN, A., & RUBIN, D.B. 1992. A single series from the Gibbs sampler provides a false sense of security. *Bayesian Statistics*, **4**, 625–631.

GELMAN, A., CARLIN, J.B., STERN, H.S., & RUBIN, D.B. 2004. *Bayesian Data Analysis, Second Edition*. Chapman and Hall/CRC, Boca Raton, USA.

GEMAN, S., & GEMAN, D. 1984. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721–741.

GEWEKE, J. 1989. Bayesian inference in econometric models using Monte Carlo integration. *Econometrica*, **57**(6), 1317–1339.

GEWEKE, J. 1992. Evaluating the accuracy of sampling-based approaches to calculation of posterior moments. *Pages 169–193 of:* BERNARDO, J.M., BERGER, J.O., DAWID, A.P., & SMITH, A.F.M. (eds), *Bayesian Statistics*, vol. 4. Oxford University Press.

GEWEKE, J., & TANIZAKI, H. 2001. Bayesian estimation of state-space models using the Metropolis-Hastings algorithm within Gibbs sampling. *Computational Statistics and Data Analysis*, **37**, 151–170.

GILKS, W.R., RICHARDSON, S., & SPIEGELHALTER, D.J. 1996. *Markov chain Monte Carlo in Practice.* Chapman and Hall / CRC, U.S.A.

GIVENS, G.H., & RAFTERY, A.E. 1996. Local adaptive importance sampling for multivariate densities with strong nonlinear relationships. *Journal of the American Statistical Association*, **91**(433), 132–141.

GOLDMAN, K.J., & ANDERSON, S.D. 1999. Space utilization and swimming depth of white sharks, *Carcharodon carcharias*, at the South Farallon islands, central California. *Environmental Biology of Fishes*, 351–364.

GOULD, J.L. 1998. Sensory bases of navigation. *Current Biology*, **8**, R731–R738.

GRAHAM, R.T., & ROBERTS, C.M. 2007. Assessing the size, growth rate and structure of a seasonal population of whale sharks (*Rhincodon typus* Smith 1828) using conventional tagging and photo identification. *Fisheries Research*, **84 Sp. Iss.**, 71–80.

GREEN, P. J. 1995. Reversible jump MCMC computation and model determination. *Biometrika*, **82**, 711–732.

GRIMM, V. 1999. Ten years of individual-based modelling in ecology: what have we learned and what could we learn in future? *Ecological Modelling*, **115**, 129–148.

GROCOTT, D.F.H. 2003. Maps in mind - how animals get home? *The Journal of Navigation*, **56**(1), 1–14.

GROENNEVIK, R., & EVENSEN, G. 2001. Application of ensemble-based techniques in fish stock assessment. *Sarsia*, **86**, 517–526.

GRUNBAUM, D. 1999. Advection-diffusion equations for generalised tactic searching behaviours. *Journal of Mathematical Biology*, **38**, 169–194.

GRUNBAUM, D. 2000. Advection-diffusion equations for internal state-mediated random walks. *SIAM Journal on Applied Mathematics*, **61**, 43–73.

GUDMUNDSSON, G. 1994. Time series analysis of catch-at-age observations. *Applied Statistics*, **43**, 117–126.

GUDMUNDSSON, G. 1995. Time series analysis of catch-at-length data. *ICES Journal of Marine Science*, **52**, 781–795.

GUILFORD, T., ROBERTS, S., BIRO, D., & REZEK, I. 2004. Positional entropy during pigeon homing II: navigational interpretation of Bayesian latent state models. *Journal of Theoretical Biology*, **227**, 25–38.

HAARIO, H., LAINE, M., MIRA, A., & SAKSMAN, E. 2006. DRAM: efficient adaptive MCMC. *Statistics and Computing*, **16**, 339–354.

HAAS, H.L., ROSE, K.A., FRY, B., MINELLO, T.J., & ROZAS, L.P. 2004. Brown shrimp on the edge: linking habitat to survival using an individual-based simulation model. *Ecological Applications*, **14(4)**, 1232–1247.

HAN, C., & CARLIN, B.P. 2001. Markov chain Monte Carlo methods for computing Bayes factors: a comparative review. *Journal of the American Statistical Association*, **96**(455), 1122–1132.

HARRISON, P.J., BUCKLAND, S.T., THOMAS, L., HARRIS, R., POMEROY, P.P., & HARWOOD, J. 2006. Incorporating movements into models of grey seal population dynamics. *Journal of Animal Ecology*, **75**, 634–645.

HARVEY, A.C. 1990. *Forecasting, Structural Time Series Models and the Kalman filter.* University of Cambridge Press, Cambridge.

HARVILLE, D.A. 1997. *Matrix Algebra from a Statistician's Perspective.* Springer-Verlag, New York.

HASTINGS, W.K. 1970. Monte Carlo sampling methods using Markov chain and their applications. *Biometrika*, **57**, 97–109.

HEIDELBERGER, P., & WELCH, P.D. 1983. Simulation run length control in the presence of an initial transient. *Operations Research*, **31**, 1109–1144.

HEUPEL, M.R., SEMMENS, J.M., & HOBDAY, A.J. 2006. Automated acoustic tracking of aquatic animals: scales, design and deployment of listening station arrays. *Marine & Freshwater Research*, **57**, 1–13.

HILBORN, R. 1990. Determination of fish movement patterns from tag recoveries using maximum likelihood estimators. *Canadian Journal of Fisheries and Aquatic Sciences*, **47**(3), 635–643.

HOBBS, N.T, & HILBORN, R. 2006. Alternatives to statistical hypothesis testing in ecology: a guide to self teaching. *Contemporary Statistics and Ecology*, **16**(1), 5–19.

HOETING, J. A., MADIGAN, D., RAFTERY, A.E., & VOLINSKY, C.T. 1999. Bayesian model averaging. *Statistical Science*, **14**, 382–401.

HORNING, M., & HILL, R.D. 2005. Designing an archival satellite transmitter for life-long deployments on oceanic vertebrates: the life history transmitter. *IEEE Transactions on Image Processing*, **30**(4), 807–817.

HOSTETLER, T.R., & KEARNEY, J.K. 2002. Strolling down the avenue with a few close friends. *Pages 7–14 of: Eurographics Ireland 2002 workshop proceedings.*

HSIAO, C.K., TZANG, J.-Y., & WANG, C.-H. 2000. Comparing the performance of two indices for spatial model selection: application to two mortality data. *Statistics in Medicine*, **19**(14), 1915–1930.

HUELSENBECK, J.P., & RONQUIST, F. 2001. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics*, **17**, 754–755.

HUELSENBECK, J.P, LARGET, B., & ALFARO, M.E. 2004. Bayesian phylogenetic model selection using reversible jump Markov chain Monte Carlo. *Molecular Biology and Evolution*, **21**(6), 1123–1133.

HUSE, G. 2001. Modelling habitat choice in fish using adapted random walk. *Sarsia*, **86(6)**, 477–483.

HUSE, G, STRAND, E, & GISKE, J. 1999. Implementing behaviour in individual-based models using neural networks and genetic algorithms. *Evolutionary Ecology*, **13**, 469–483.

JEANSON, R., BLANCO, S., FOURNIER, R., DENEUBOURG, J.-L., FOURCASSIE, V., & THERAULAZ, G. 2003. A model of animal movements in a bounded space. *Journal of Theoretical Biology*, **225**, 443–451.

JEFFREYS, H. 1935. Some tests of significance, treated by the theory of probability. *Proceedings of the Cambridge Philosophy Society*, **31**, 203–222.

JEFFREYS, H. 1961. *Theory of Probability.* Oxford University Press, Oxford, U.K.

JENNRICH, R.I., & TURNER, F.B. 1969. Measurement of noncircular home range. *Journal of Theoretical Biology*, **22**, 227–237.

JONES, R. 1976. The use of marking data in fish population analysis. *FAO Fisheries Technical Paper*, **153**, 1–42.

JONES, R.H., & ACKERSON, L.M. 1990. Serial correlation in unequally spaced longitudinal data. *Biometrika*, **77**(4), 721–731.

JONSEN, I.D., MYERS, R.A., & FLEMMING, J.E.M. 2003. Meta-analysis of animal movement using state-space models. *Ecology*, **84(11)**, 3055–3063.

JONSEN, I.D., FLEMMING, J.M., & MYERS, R.A. 2005. Robust state-space modeling of animal movement data. *Ecology*, **86**(11), 2874–2880.

JONSEN, I.D., MYERS, R., & JAMES, M.C. 2006. Robust hierarchical state-space models reveal diel variation in travel rates of migrating leatherback turtles. *Journal of Animal Ecology*, **75**, 1046–1057.

JONSEN, I.D., MYERS, R.A., & JAMES, M.C. 2007. Identifying leatherback turtle foraging behaviour from satellite telemetry using a switching state-space model. *Marine Ecology Progress Series*, **337**, 255–264.

JULIER, S.J., & UHLMANN, J.K. 2004. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, **92**, 401–422.

JULIER, S.J., UHLMANN, J.K., & DURRANT-WHYTE, H. 1995. A new approach for filtering nonlinear systems. *Proceedings of the American Control Conference*, 1628–1631.

KADANE, J.B., & LAZAR, N.A. 2004. Methods and criteria for model selection. *Journal of the American Statistical Association*, **99**(465), 279–290.

KALMAN, R.E. 1960. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering, Transactions ASME, Series D*, **82**, 35–45.

KALMAN, R.E., & BUCY, R.S. 1961. New results in linear filtering and prediction theory. *Journal of Basic Engineering, Transactions ASME, Series D*, **83**, 95–108.

KAREIVA, P.M., & SHIGESADA, N. 1983. Analysing insect movement as a correlated random walk. *Oecologia*, **56**, 234–238.

KASAI, A., KISHI, M.J., & SUGIMOTO, T. 1992. Modeling the transport and survival of Japanese sardine larvae in and around the Kuroshio current. *Fisheries Oceanography*, **1**(1), 1–10.

KASS, R.E. 1993. Bayes factors in practice. *The Statistician*, **42**, 551–560.

KASS, R.E., & RAFTERY, A.E. 1995. Bayes factors. *Journal of the American Statistical Association*, **90**(430), 773–795.

KENWARD, R. 2001. *Wildlife Radio Tagging: Equipment, Field Techniques and Data Analysis.* London: Academic Press.

KIM, C.-J., & NELSON, C.R. 1999. *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications.* The MIT Press, Cambridge, Massachusetts, U.S.A.

KIMURA, D.K., BALSINGER, J.W., & ITO, D.H. 1996. Kalman filtering the delay-difference equation: practical approaches and simulations. *Fishery Bulletin*, **94**, 678–691.

KITIGAWA, G. 1987. Non-Gaussian state-space modeling of nonstationary time series (with discussion). *Journal of the American Statistical Association*, **82**, 1032–1063.

KLAFTER, J., SHLESINGER, M.F., & ZUMOFEN, G. 1996. Beyond Brownian motion. *Physics Today*, **49**(2), 33–39.

KLIMLEY, A.P. 1993. Highly directional swimming by scalloped hammerhead sharks, *Sphyrnà lewini*, and subsurface irradiance, temperature, bathymetry, and geomagnetic field. *Marine Biology*, **117**(1), 1–22.

KLIMLEY, A.P., BEAVER, S.C., CURTIS, T.H., & JORGENSEN, S.J. 2002. Movements and swimming behaviour of three species of sharks in La Jolla canyon, California. *Environmental Biology of Fishes*, **63**(2), 117–135.

KLOEK, T., & VAN DIJK, H.K. 1978. Bayesian estimates of equation system parameters: an application of integration by Monte Carlo. *Econometrica*, **46**, 1–20.

KOHLER, N.E., & TURNER, P.A. 2001. Shark tagging: a review of conventional methods and studies. *Environmental Biology of Fishes*, **60**(1-3), 191–223.

LARKIN, R.P., & HALKIN, D. 1994. A review of software packages for estimating animal home ranges. *Wildlife Society Bulletin*, **22**, 274–287.

LAVINE, M., & SCHERVISH, M.J. 1999. Bayes factors: what they are and what they are not. *The American Statistician*, **53**(2), 119–122.

LAWSON, E.J.G., & RODGERS, A.R. 1997. Differences in home-range size computed in commonly used software packages. *Wildlife Society Bulletin*, **25(3)**, 721–729.

LEE, S-.Y., & SONG, X-.Y. 2004. Bayesian model comparisons of nonlinear structural equation models with missing continuous and ordinal categorical data. *British Journal of Mathematical and Statistical Psychology*, **57**, 131–150.

LEE, S.-Y., & TANG, N.-S. 2006. Bayesian analysis of nonlinear structural equation models with nonignorable missing data. *Psychometrika*, **71**(3), 541–564.

LEIS, J.M., & CARSON-EWART, B.M. 2003. Orientation of pelagic larvae of coral-reef fishes in the ocean. *Marine Ecology Progress Series*, **252**, 239–253.

LENTFER, J.W., & DEMASTER, D. 1982. Satellite radio tracking of polar bears. NASA SP-457, Washington, D.C.

LEUZE, C.C.K. 1980. A handbook on biotelemetry and radio tracking. Pergamon Press, Oxford, England.

LEVANDOWSKY, M., WHITE, B.S., & SCHUSTER, F. 1997. Random movements of soil amoebas. *Acta Protozoologica*, **36**, 237–248.

LEWIS, S.M., & RAFTERY, A.E. 1997. Estimating Bayes factors via posterior simulation with the Laplace-Metropolis estimator. *Journal of the American Statistical Association*, **92**(438), 648–655.

LINDLEY, D.V. 1957. A statistical paradox. *Biometrika*, **44**, 187–192.

LIU, J., & WEST, M. 2001. Sequential Monte Carlo methods in practice. Springer-Verlag, New York.

LIU, J.S. 2004. *Monte Carlo Strategies in Scientific Computing*. Spring Science + Business Media, LLC, New York, USA.

LIU, J.S., & CHEN, R. 1998. Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statististical Association*, **93**, 1032–1044.

LOHMANN, K.J, & LOHMANN, C.M.F. 2006. Sea turtles, lobsters, and oceanic magnetic maps. *Marine and Freshwater Behaviour and Physiology*, **39**(1), 49–64.

LOPES, H.F., & WEST, M. 2004. Bayesian model assessment in factor analysis. *Statistica Sinica*, **14**, 41–67.

LOWRY, L.F, FROST, K.J., VER HOEF, J.M., & DEJONG, R.A. 2001. Movements of satellite-tagged subadult and adult harbor seals in Prince William Sound, Alaska. *Marine Mammal Science*, **17**(4), 835–861.

MACDONALD, I.L., & ZUCCHINI, W. 1997. *Hidden Markov and other Models for Discrete-Valued Time Series.* Chapman and Hall / CRC, Boca Raton, USA.

MANDELBROT, B. B. 1983. *The Fractal Geometry of Nature.* W. H. Freeman, San Francisco, California, USA.

MANNING, A., & DAWKINS, M.S. 1992. *An Introduction to Animal Behaviour.* Cambridge University Press.

MATIS, J.H., GRANT, W.E., & MILLER, T.H. 1992. A semi-Markov process model for migration of marine shrimp. *Ecological Modelling*, **60**, 167–184.

MATTHIOPOULOS, J. 2003. Model-supervised kernel smoothing for the estimation of spatial usage. *OIKOS*, **102**, 367–377.

MATTHIOPOULOS, J. 2005. Metapopulation consequences of site fidelity for colonially breeding mammals and birds. *Journal of Animal Ecology*, **74**, 716–727.

MAURITZEN, M., BELIKOV, S.E., BOLTUNOV, A.N., DEROCHER, A.E., HANSEN, E., IMS, R.A., WIIG, O., & YOCCOZ, N. 2003. Functional responses in polar bear habitat selection. *OIKOS*, **100(1)**, 112–124.

MAURY, O., & GASCUEL, D. 1999. SHADYS ('simulateur halieutique de dynamiques spatiales') a GIS based numerical model of fisheries. Example application: the study of a marine protected area. *Aquatic Living Resources*, **12**(2), 77–88.

MAURY, O., GASCUEL, D., & FONTENEAU, A. 2001. Spatial modeling of Atlantic yellowfin tuna population dynamics: Application of a habitat-based advection-diffusion-reaction model to the study of local overfishing. Spatial Processes and Management of Marine Populations. University of Alaska, Alaska Sea Grant College program, AK-SG-01-02.

MAYERS, D. F., & MORTON, K. W. 2005. *Numerical Solution of Partial Differential Equations: An Introduction.* Cambridge University Press, Mathematics.

MCALLISTER, M.K., & IANELLI, J.N. 1997. Bayesian stock assessment using catch-age data and the sampling-importance resampling algorithm. *Canadian Journal of Fisheries and Aquatic Sciences*, **54**, 284–300.

MCCONNELL, B.J., FEDAK, M.A., LOVELL, P., & HAMMOND, P. 1999. Movement and foraging areas of grey seals in the North Sea. *Journal of Applied Ecology*, **36**, 573–590.

MCDONALD, A.D., & SANDAL, L.K. 1999. Estimating the parameters of stochastic differential equations using a criterion function based on the Kolmogarov-Smirnov statistic. *Journal of Statistical Computation and Simulation*, **64**, 235–250.

MECH, L. D., & BARBER, S. M. 2002. *A critique of wildlife radio-tracking and its use in national parks: a report to the U.S. National Park Service.* Tech. rept. U.S. Geological Survey, Northern Prairie Wildlife Research Center, Jamestown, N.D. Jamestown, ND: Northern Prairie Wildlife Research Center Online. http://www.npwrc.usgs.gov/resource/wildlife/radiotrk/index.htm (Version 30DEC2002).

MEINHOLD, R.J., & SINGPURWALLA, N.D. 1983. Understanding the Kalman filter. *The American Statistician*, **37(2)**, 123–127.

MENG, X.-L., & WONG, M.H. 1996. Simulating ratios of normalising constants via a simple identity: a theoretical exploration. *Statistica Sinica*, **6**, 831–860.

METROPOLIS, N., ROSENBLUTH, A.W., ROSENBLUTH, M.N., TELLER, A.H., & TELLER, E. 1953. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, **21**, 1087–1091.

MEYER, R., & MILLAR, R.B. 1999a. Bayesian stock assessment using a state-space implementation of the delay difference model. *Canadian Journal of Fisheries and Aquatic Sciences*, **56**(1), 37–52.

MEYER, R., & MILLAR, R.B. 1999b. BUGS in Bayesian stock assessments. *Canadian Journal of Fisheries and Aquatic Sciences*, **56**, 1078–1086.

MILLAR, R.B. 2004. Simulated maximum likelihood applied to non-Gaussian and nonlinear mixed effects and state-space models. *Australian and New Zealand Journal of Statistics*, **46(4)**, 543–554.

MILLAR, R.B., & MEYER, R. 2000a. Bayesian state-space modeling of age-structured data: fitting a model is just the beginning. *Canadian Journal of Fisheries and Aquatic Sciences*, **57**, 43–50.

MILLAR, R.B., & MEYER, R. 2000b. Non-linear state space modelling of fisheries biomass dynamics by using Metropolis-Hastings within-Gibbs sampling. *Applied Statistics*, **49(3)**, 327–342.

MONTGOMERY, J.C., & WALKER, M.M. 2001. Orientation and navigation in elasmobranchs: Which way forward? *Environmental Biology of Fishes*, **60**(1-3), 109–116.

MORALES, J.M., HAYDON, D.T., FRIAR, J., HOLSINGER, K.E., & FRYXELL, J.M. 2004. Extracting more out of relocation data: building movement models as mixtures of random walks. *Ecology*, **85(9)**, 2436–2445.

MULLEN, A.J. 1989. Aggregation of fish through variable diffusivity. *Fishery Bulletin*, **87**, 353–362.

MULLER, K.-R., MIKA, S., RATSCH, G., TSUDA, K., & SCHOLKOPF, B. 2001. An introduction to kernel-based learning algorithms. *IEEE Transactions on Neural Networks*, **12(2)**, 181–202.

MURPHY, E.J., THORPE, S.E., WATKINS, J.L., & HEWITT, R. 2004. Modeling the krill transport pathways in the Scotia sea: spatial and environmental connections generating the seasonal distribution of krill. *Deep-Sea Research II*, 1435–1456.

MURRAY, J.D. 1989. *Mathematical Biology.* Springer-Verlag, Heidelberg.

MYERS, R.A., HAMMILL, M.O., & STENSON, G.B. 1997. Using mark-recapture to estimate the numbers of a migrating stage-structured population. *Canadian Journal of Fisheries and Aquatic Sciences*, **54**, 2097–2104.

NEAL, R.M. 2001. Annealed importance sampling. *Statistics and Computing*, **11**, 125–139.

NEWMAN, K.B. 1998. State-space modeling of animal movement and mortality with application to salmon. *Biometrics*, **54**(4), 1290–1314.

NEWMAN, K.B. 2000. Hierarchical modelling of salmon harvest and migration. *Journal of Agricultural, Biological and Environmental Statistics*, **5**(4), 430–455.

NEWMAN, K.B., & LINDLEY, S.T. 2006. Accounting for demographic and environmental stochasticity, observation error, and parameter uncertainty in fish population dynamics models. *North American Journal of Fisheries Management*, **26**, 685–701.

NEWTON, M.A., & RAFTERY, A.E. 1994. Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society B*, **56**(1), 3–48.

OKE, P.R., SCHILLER, A., GRIFFIN, D.A., & BRASSINGTON, G.B. 2005. Ensemble data assimilation for an eddy-resolving ocean model of the Australian region. *Quarterly Journal of Applied Meteorology, Part C*, **131 (613)**, 3301–3311.

OKUBO, A. 1980. *Diffusion and Ecological Problems: Mathematical Models.* Vol. 10. Springer-Verlag, New York.

OKUBO, A., & LEVIN, S.A. 2001. *Diffusion and Ecological Problems, Modern Perspectives.* Springer-Verlag, New York.

OLLASON, J.G., BRYANT, A.D., DAVIS, P.M., SCOTT, B.E., & TASKER, M.L. 1997. Predicted seabird distributions in the North Sea: the consequences of being hungry. *Seabirds in the Marine Environment, Proceedings of an ICES International Symposium held in Glasgow*, **54**(4), 507–517.

OSTFELD, R.S. 1982. Foraging strategies and prey switching in the California sea otter. *Oecologia (Berlin)*, **53**, 170–178.

OVASKAINEN, O. 2004. Habitat-specific movement parameters estimated using mark-recapture data and a diffusion model. *Ecology*, **85(1)**, 242–257.

PALMER, J.B. 1992. Hierarchical and concurrent individual based modelling. *Pages 188–207 of:* DEANGELIS, D.L., & GROSS, L.J (eds), *Individual-based models and approaches in ecology: populations, communities and ecosystems.* Chapman and Hall, New York.

PARDINI, A.T., JONES, C.S., NOBLE, L.R., KREISER, B., MALCOLM, H., BRUCE, B.D., STEVENS, J.D., CLIFF, G., SCHOLL, M.C., FRANCIS, M., DUFFY, C.A.J., & MARTIN, A.P. 2001. Sex-biased dispersal of great white sharks. *Nature*, **412**, 139–140.

PARMALEE, D.F., PARMALEE, J.M., & FULLER, M. 1985. Ornithological investigations at Palmer station: the first long-distance tracking of seabirds by satellites. *Antarctic Journal of the United States*, **19**, 162–163.

PASCUAL, M.A., & QUINN, T.P. 1991. Evaluation of alternative models of the coastal migration of adult Fraser river salmon (*Oncorhynchus nerka*). *Canadian Journal of Fisheries and Aquatic Sciences*, **48**, 799–810.

PATTERSON, T.A., THOMAS, L., WILCOX, C., OVASKAINEN, O., & MATTHIOPOULOS, J. 2007. State-space models of individual animal movement. *Trends in Ecology & Evolution*, **894**, 1–8.

PAWITAN, Y. 2001. *In all Likelihood: Statistical Modelling and Inference using Likelihood.* Oxford Science Publications, U.K.

PELLA, J.J. 1993. Utility of structural time series models and the Kalman filter for predicting consequences of fishery actions. *In:* KRUSE, G., EGGERS, D.M., MARASCO, R.J., PAUTZKE, C., & QUINN, T.J. II (eds), *Proceedings of the International Symposium on Management Strategies for Exploited Fish Populations.* Alaska Sea Grant College Program Report No. 93-02, University of Alaska, Fairbanks, Alaska, p 571-593.

PETKOVIC, P., & BUCHANAN, C. 2002. *Australian bathymetry and topography grid. [CDROM].* Tech. rept. Canberra: Geoscience Australia.

POLACHECK, T. 1990. Year round closed areas as a management tool. *Natural Resource Modelling*, **4(3)**, 327–354.

PORCH, C.E. 1995. Trajectory-based approaches to estimating velocity and diffusion from tagging data. *Fishery Bulletin*, **93**, 694–709.

PREISLER, H.K., AGER, A.A., JOHNSON, B.K., & KIE, J.G. 2004. Modeling animal movements using stochastic differential equations. *Environmetrics*, **15**, 643–657.

PRIEDE, I.G., & FRENCH, J. 1991. Tracking of marine animals by satellite. *International Journal of Remote Sensing*, **12**(4), 667–680.

PUNT, A.E. 2003. Extending production models to include process error in the population dynamics. *Canadian Journal of Fisheries and Aquatic Sciences*, **60**(10), 1217–1228.

QUINN, T. 1994. How do sharks orient at sea? *Trends in Ecology & Evolution*, **9(8)**, 277–278.

R DEVELOPMENT CORE TEAM. 2007. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

RABINER, L.R. 1989. A tutorial on hidden Markov models and selected application in speech recognition. *Proceedings of the IEEE*, **77**(2), 257–286.

RAFTERY, A.E. 1996. Approximate Bayes factors and accounting for model uncertainty in generalised linear models. *Biometrika*, **83**(2), 251–266.

RAFTERY, A.E., & LEWIS, S.M. 1992. How many iterations in the Gibbs sampler? *Pages 763–773 of:* BERNARDO, J.M., BERGER, J.O., DAWID, A.P., & SMITH, A.F.M. (eds), *Bayesian statistics 4*. Oxford University Press, Oxford, U.K.

RAILSBACK, S.F. 2001. Concepts from complex adaptive systems as a framework for individual-based modelling. *Ecological Modelling*, **139**, 47–62.

RAILSBACK, S.F., STAUFFER, H.B., & HARVEY, B.C. 2003. What can habitat preference models tell us? Tests using a virtual trout population? *Ecological Applications*, **13(6)**, 1580–1594.

REED, W.J., & SIMONS, C.M. 1996. Analysing catch-effort data by means of the Kalman filter. *Canadian Journal of Fisheries and Aquatic Sciences*, **53**, 2157–2166.

RIVOT, E., PREVOST, E., PARENT, E., & BAGLINIERE, J.L. 2004. A Bayesian state-space modelling framework for fitting salmon stage-structured population dynamic model to multiple time series of field data. *Ecological Modelling*, **179**(4), 463–485.

ROBERT, C.P., & CASELLA, G. 1999. *Monte Carlo Statistical Methods.* Springer Science + Business Media, LLC, U.S.A.

ROBERTS, G.O. 1992. Convergence diagnostics of the Gibbs sampler. *Pages 775–782 of:* BERNARDO, J.M., BERGER, J.O., DAWID, A.P., & SMITH, A.F.M. (eds), *Bayesian Statistics 4*. Oxford University Press, Oxford.

ROBERTS, S., GUILFORD, T., REZEK, I, & BIRO, D. 2004. Position entropy during pigeon homing I: application of Bayesian latent stage modelling. *Journal of Theoretical Biology*, **227**, 39–50.

RONQUIST, F., & HUELSENBECK, J.P. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.

ROOT, R.B., & KAREIVA, P.M. 1984. The search for resources by cabbage butterflies (*Pieres rapae*): ecological consequences and adaptive significance of Markovian moments in a patchy environment. *Ecology*, **65**(1), 147–165.

ROSENKRANTZ, S.L., & RAFTERY, A.E. 1994. *Covariate selection in hierarchical models of hospital admission counts: a Bayes factor approach.* Tech. rept. Department of Statistics, GN-22, University of Washington, Seattle, Washington, 98195 USA.

ROYER, F., FROMENTIN, J.M., & GASPAR, P. 2005. A state-space model to derive bluefin tuna movement and habitat from archival tags. *OIKOS*, **109**(3), 473–484.

RUBIN, D.B. 1987. Comment on 'The calculation of posterior distributions by data augmentation.'. *Journal of the American Statistical Association*, **82**, 543–546.

RUBIN, D.B. 1988. Using the SIR algorithm to simulate posterior distributions. *Pages 385–402 of:* BERNARDO, J.M., DEGROOT, M.H., LINDLEY, D.V., & SMITH, A.M. (eds), *Bayesian statistics 3: Proceedings of the third Valencia international meeting, June 1-5, 1987.* Clarendon Press, Oxford.

RYERSON, R.A. 1998. *Manual of Remote Sensing.* John Wiley & Sons Ltd, New York.

SCHNUTE, J.T. 1991. The importance of noise in fish population models. *Fisheries Research*, **11**, 197–223.

SCHNUTE, J.T. 1994. A general framework for developing sequential fisheries models. *Canadian Journal of Fisheries and Aquatic Sciences*, **51**, 1676–1688.

SCHWARZ, G. 1978. Estimating the dimension of a model. *Annals of Statistics*, **6**, 461–464.

SCHWEPPE, F. 1965. Evaluation of likelihood functions for Gaussian signals. *IEEE Transactions on Information Theory*, **11**, 61–70.

SHLESINGER, M.F. 2001. Physics in the noise. *Nature*, **411**, 641.

SHUTLER, J.D., SMYTH, T.J., LAND, P.E., & GROOM, S.B. 2005. A near-real time automatic MODIS data processing system. *International Journal of Remote Sensing*, **26(5)**, 1049–1055.

SIBERT, J.R. 1984. A two-fishery tag attrition model for the analysis of mortality, recruitment and fishery interaction. *Tuna and Billfish Assessment Programme Technical Report*, **13**, 1–27.

SIBERT, J.R., & FOURNIER, D.A. 1994. Evaluation of advection-diffusion equations for estimation of movement patterns from tag recapture data. *FAO Fisheries Technical Paper*, **336/1**, 108–121.

SIBERT, J.R., HAMPTON, J., & FOURNIER, D.A. 1996. Skipjack movement and fisheries interaction in the western Pacific. *FAO Fisheries Technical Paper*, **365**, 402–418.

SIBERT, J.R., HAMPTON, J., FOURNIER, D.A., & BILLS, P.J. 1999. An advection-diffusion-reaction model for the estimation of fish movement parameters from tagging data, with application to skipjack tuna (*Katsuwonus pelamis*). *Canadian Journal of Fisheries and Aquatic Sciences*, **56**(6), 925–938.

SIBERT, J.R., MUSYL, M.K., & BRILL, R.W. 2003. Horizontal movements of bigeye tuna (*Thunnus obesus*) near Hawaii determined by Kalman filter analysis of archival tagging data. *Fisheries Oceanography*, **12(3)**, 141–151.

SIBERT, J.R., LUTCAVAGE, M.E., NIELSEN, A., BRILL, R.W., & WILSON, S.G. 2006. Interannual variation in large-scale movement of Atlantic bluefin tuna (*Thunnus thynnus*) determined from pop-up satellite archival tags. *Canadian Journal of Fisheries and Aquatic Sciences*, **63**, 2154–2166.

SINNOTT, R. W. 2004. Virtues of the Haversine. *Sky and Telescope*, **68 (2)**, 159.

SKELLAM, J.G. 1951. Random dispersal in theoretical populations. *Biometrika*, **38**, 196–218.

SORENSEN, D., & GIANOLA, D. 2002. *Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics.* Springer Science+Business Media, Inc., USA.

SPIEGELHALTER, D.J., BEST, N., & CARLIN, B.P. 1998. *Bayesian deviance, the effective number of parameters, and the comparison of arbitrarily complex models.* Tech. rept. 98-009. Division of Biostatistics, University of Minnesota.

SPIEGELHALTER, D.J., BEST, N.G., CARLIN, B.P., & VAN DER LINDE, A. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B*, **64**(4), 583–639.

SPIEGELHALTER, D.J., THOMAS, A., BEST, N.G., & LUNN, D. 2004. *Win-BUGS User Manual, version 2.0.* Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge.

STEWART, L. 1983. Bayesian analysis using Monte Carlo integration - a powerful methodology for handling some difficult problems. *The Statistician*, **32**(1/2), 195–200.

STRONG, W.R. JR, MURPHY, R.C., BRUCE, B.D., & NELSON, D.R. 1992. Movements and associated observations of bait-attracted white sharks, *Carcharodon carcharias*: A preliminary report. *Australian Journal of Marine & Freshwater Research*, **43(1)**, 13–20.

STRONG, W.R. JR., BRUCE, B.D., NELSON, D.R., & MURPHY, R.D. 1996. Great white sharks: The biology of *Carcharodon carcharias*. Academic Press, San Diego.

SULLIVAN, P.J. 1992. A Kalman filter approach to catch-at-age analysis. *Biometrics*, **48**, 237–257.

SUMNER, M., WOTHERSPOON, S.J., & HINDELL, M. *In Review for Ecology. A general framework for the estimation of animal movement from archival and telemetry techniques.*

SUNDSTROM, L.F., GRUBER, S.H., CLERMONT, S.M., CORREIA, J.P.S., DE MARIGNAC, J.R.C., MIRRISSEY, J.F., LOWRANCE, C.R., THOMASSEN, L., & OLIVEIRA, M.T. 2001. Review of elasmobranch behavioural studies using ultrasonic telemetry with special reference to the lemon shark, *Negaprion brevirostris*, around Bihmini islands, Bahamas. *Environmental Biology of Fishes*, **60**, 225–250.

TANNER, M.A., & WONG, W.H. 1987. The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, **82**(398), 528–540.

TEO, S.L.H., BOUSTANY, A., BLACKWELL, S., WALLI, A., WENG, K.C., & BLOCK, B.A. 2004. Validation of geolocation estimates based on light level and sea surface temperature from electronic tags. *Marine Ecology Progress Series*, **283**, 81–98.

THOMAS, L., BUCKLAND, S.T., NEWMAN, K.B., & HARWOOD, J. 2005. A unified framework for modelling wildlife population dynamics. *Australian and New Zealand Journal of Statistics*, **47**(1), 19–34.

THOMPSON, D., LONERGAN, M., & DUCK, C. 2005. Population dynamics of harbour seals *Phoca vitulina* in England: monitoring growth and catastrophic declines. *Journal of Applied Ecology*, **42**(4), 638–648.

TIERNEY, L. 1994. Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics*, **22**, 1701–1762.

TIERNEY, L., & KALDANE, J.B. 1986. Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, **81**, 82–86.

TREMBLAY, Y., SHAFFER, S.A., FOWLER, S.L., KUHN, C.E., MCDONALD, B.I., WEISE, M.J., BOST, C.-A., WEIMERSKIRCH, H., CROCKER, D.E., GOEBEL, M.E., & COSTA, D.P. 2006. Interpolation of animal tracking data in a fluid environment. *The Journal of Experimental Biology*, **209**, 128–140.

TRENKEL, V.M., ELSTON, D.A., & BUCKLAND, S.T. 2000. Fitting population dynamics models to count and cull data using sequential importance sampling. *American Statistical Association*, **95 (450)**, 363–374.

TURCHIN, P. 1996. Fractal analysis of animal movement: a critique. *Ecology*, **77(7)**, 2086–2090.

TURCHIN, P. 1998. *Quantitative Analysis of Movement.* Sunderland, Massachusetts: Sinauer Associates Inc.

VACHHANI, P., NARASIMHAN, S., & RENGASWAMY, R. 2006. Robust and reliable estimation via unscented recursive nonlinear dynamic data reconciliation. *Journal of Process Control*, **16**, 1075–1086.

VAN VUREN, D. 1998. Behavioral ecology and conservation biology. Oxford University Press, U.K.

VINCENT, C., MCMONNELL, B.J., RIDOUX, V., & FEDAK, M.A. 2002. Assessment of Argos location accuracy from satellite tags deployed on captive gray seals. *Marine Mammal Science*, **18**, 156–166.

VISWANATHAN, G.M., AFANASYEV, V., BULDYREV, S.V., MURPHY, E.J., PRINCE, P.A., & STANLEY, H.E. 1996. Levy flight search patterns of wandering albatross. *Nature*, **381**(6581), 413–415.

VISWANATHAN, G.M., BULDYREV, S.V., HAVLIN, S., DA LUZ, M.G.E., RAPASO, E.P., & STANLEY, H.E. 1999. Optimizing the success of random searches. *Nature*, **401**, 911–914.

VOIGHT, D.R., & TINLINE, R.R. 1980. Strategies for analysing radio tracking data. *Pages 387–404 of:* AMLANER, C.J., & MACDONALD, D.W. (eds), *A Handbook on Biotelemetry and Radio Tracking.* Oxford: Pergamon Press.

WALKER, D.M., PEREZ-BARBERIA, F.J., & MARION, G. 2006. Stochastic modelling of ecological processes using hybrid Gibbs samplers. *Ecological Modelling*, **198**, 40–52.

WALKER, M.M., DENNIS, T.E., & KIRSCHVINK, J.L. 2002. The magnetic sense and its use in long-distance navigation by animals. *Current Opinion in Neurobiology*, **12**(6), 735–744.

WALLRAFF, H.G. 2004. Avian olfactory navigation: its empirical foundation and conceptual state. *Animal Behaviour*, **67**(2), 189–204.

WALTERS, C.J., HANNAH, C.G., & THOMSON, K. 1992. A microcomputer program for stimulating effects of physical transport processes on fish larvae. *Fisheries Oceanography*, **1**, 11–19.

WANG, F., & MCKENZIE, E. 1999. A multi-agent evolutionary artificial neural network for general navigation in unknown environments. *Proceedings of the Third International Conference on Autonomous Agents, Seattle*, 154–159.

WANG, G. 2007. On the latent state estimation of nonlinear population dynamics using Bayesian and non-Bayesian state-space models. *Ecological Modelling*, **200**, 521–528.

WASER, P.M. 1976. *Cercocebus albigena*: site attachment, avoidance and intergroup spacing. *American Naturalist*, **110**, 911–935.

WENG, K.C., BOUSTANY, A.M., PYLE, P., ANDERSON, S.D., BROWN, A., & BLOCK, B.A. 2007. Migration and habitat of white sharks (*Carcharodon carcharias*) in the eastern Pacific ocean. *Marine Biology*, **152**, 877–894.

WERNER, F.E., QUINLAN, J.A., LOUGH, R.G., & LYNCH, D.R. 2001. Spatially-explicit individual based modeling of marine populations: a review of the advances in the 1990s. *Sarsia*, **86**(6), 411–421.

WEST, M. 1993. Approximating posterior distributions by mixtures. *Journal of the Royal Statistical Society Series B*, **55**, 409–422.

WEST, M., & HARRISON, J. 1997. *Bayesian Forecasting and Dynamic Models.* New York: Springer-Verlag.

WHITE, G.C., & GARROTT, R.A. 2006. *Analysis of Wildlife Tracking Data.* Academic Press Ltd, San Diego, California.

WIKTORSSON, M., RYDEN, T., NILSSON, E., & BENGTSSON, B. 2004. Modelling the movement of a soil insect. *Journal of Theoretical Biology*, **231**(4), 497–513.

WOOD, A.G., NAEF-DAENZER, B., PRINCE, P.A., & CROXALL, J.P. 2000. Quantifying habitat use in satellite-tracked pelagic seabirds: application of kernel estimation to albatross locations. *Journal of Avian Biology*, **31**, 278–286.

WORTON, B.J. 1989. Kernel methods for estimating the utilization distribution in home-range studies. *Ecology*, **70(1)**, 164–168.

XIAO, Y. 2000. An individual-based approach to evaluating experimental designs for estimating rates of fish movement from tag recoveries. *Ecological Modelling*, **128**, 149–163.

XIAOHUA, D., SHANNON, G., SLOTOW, R., PAGE, B., & DUFFY, K.J. 2007. Short-duration daytime movements of a cow herd of African elephants. *Journal of Mammology*, **88**, 151–157.

YAMAMURA, K., MORIYA, S., & TANAKA, K. 2003. Discrete random walk model to interpret the dispersal parameters of organisms. *Ecological Modelling*, **161**, 151–157.

ZOLLNER, P.A., & LIMA, S.L. 1999. Search strategies for landscape-level interpatch movements. *Ecology*, **80**(3), 1019–1030.